FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Md. Surat-E- Mostafa

Estimating Spectral HRV Features with Missing Data

Master's Thesis
Degree Programme in Computer Science and Engineering
December 2018

Mostafa M. (2018) Estimating Spectral HRV Features with Missing Data. University of Oulu, Degree Programme in Computer Science and Engineering. Master's thesis, 59 p.

ABSTRACT

Physiological signals, ECG signal, have been widely used for diagnosis, disease identification and nowadays for self-monitoring. Missing data represents the problem in spectral analysis. This study focuses on the HRV power spectral analysis in frequency-domain using three methods with simulated missing data in real RR interval tachograms. Actual missing ECG data are collected from the unconstrained measurement. Parametric, Non-parametric and uneven sampling approach were used for calculating the power spectral density (PSD), and cubic spline interpolation method was used for the non-parametric method. Based on this studies outcome, the effect of missing R-R interval data and optimal method was observed through the simulated real R-R interval tachograms for missing data. About 0 to 6 percentage data were removed according to the exponential Poisson distribution from the real R-R interval data for normal sinus rhythm, atrial fibrillation, tachycardia and bradycardia patient which data obtained from MIT-BIH Arrhythmia database to simulate real-world packet loss. For this analysis, 5 min duration data were used in all and 1000 Monte Carlo runs is performed for certain percentage missing data. Power spectral density (PSD) corresponding each frequency component was estimated as the frequency-domain parameters in each run and error power percentage based on each element difference between with and without the missing data duration were calculated. In addition, this study revealed that power spectral entropy measurement from power spectral density which differentiates between different arrhythmias.

Keywords: ECG, heart rate variability (HRV), frequency domain, power spectral density (PSD), power spectral entropy, missing data, RRI, FFT method, AR model, LS method, arrhythmias.

TABLE OF CONTENTS

FOREWORD

From the beginning of my student life, I have a dream to learn new thing, to build my future career. For making my dreams in reality, I have been started my journey at the University of Oulu as a Master's student in Biomedical Engineering: Signal and Image processing. I would like to respect my parents because they always support me to fulfill my dreams.

Since the beginning of my study, my supervisor, Professor Tapio Seppänen gave me a chance to work in his research group. It was the first job in my life as a research assistant. A secondary phase of the project that I have been participating in was dedicated to my Master's thesis. This research work aims to calculate robust spectral HRV features with missing data.

I am obliged to give my gratitude to my advisor Professor Tapio Seppänen for giving me an opportunity to hold this research work. I would also like to express the gratitude my technical advisor Kai Noponen. During my job here at the Center for Machine Vision and Signal Analysis, I have been technically guided by Kai Noponen, and Professor Tapio Seppänen advised me regarding my thesis work. The discussions, encouragement and constant feedback from Professor Tapio Seppänen and Kai Noponen taught me how to address a research hypothesis efficiently. With this, I would like to thank my examiners, advisers, and supervisor.

University supervisor: Dr. Sc. (Tech.) Prof. Tapio Seppänen
Second examiner: Dr. Sc. (Tech.) Jukka Kortelainen
Technical supervisor: M.Sc. Kai Noponen

LIST OF ABBREVIATIONS AND SYMBOLS

WOSA	Weighted Overlapped Segment Averaging
PMF	Probability Mass Function
PSD	Power Spectral Density
ECG	Electrocardiogram
EEG	Electroencephalogram
SA	Sinoatrial
BPM	Beats Per Minute
AF	Atrial Fibrillation
HRV	Heart Rate Variability
ANS	Autonomic Nervous System
FDA	Food and Drug Administration
FFT	Fast Fourier Transform
AR	Autoregressive Model
LS	Lomb-Scargle
HF	High-Frequency
LF	Low-Frequency
WEE	Wavelet Energy Entropy
WTE	Wavelet Time Entropy
WSE	Wavelet Singular Entropy
WTFE	Wavelet Time-Frequency Entropy
WAE	Wavelet Average Entropy
WDE	Wavelet Distance Entropy
LF	Low-Frequency
CHF	Congestive Heart Failure
RRI	R-R Interval
SA	Sinoatrial
NSR	Normal Sinus Rhythm
PSE	Power Spectrum Entropy
PDF	Probability Density Function
VT	Ventricular Tachycardia
VF	Ventricular Fibrillation
SNS	Sympathetic Nervous System
IIR	Infinite Impulse Response
NaN	Not-a-Number
REs	Relative Errors
RMSRE	Root-Mean-Squared Relative Error
HR	Heart Rate
AV	Atrioventricular
RA	Right Atrium
LA	Left Atrium
RV	Right Ventricle
LV	Left Ventricle
STFT	Short Time Fourier Transform
PNS	Parasympathetic Nervous System
ACF	Auto-Correlation Function

SNR	Signal to Noise Ratio
HR	Heart Rate
DTF	Discrete Time Frequency

1. INTRODUCTION

The analysis of heart rate variability is a process to study the automatic nervous system (ANS) and the evaluation of clinical cardiac risk. To observe the heart rate variability is performed in the time and frequency domain techniques among the several proposed techniques [1]. Spectral analysis was first introduced in astronomical time-series analysis for unevenly sampled data in 1969. Besides, Task Force of the European Society of Cardiology in the North American described the HRV estimation methods in 1996 [2]. After that, the US Food and Drug Administration (FDA) approved the medical devices for calculating the HRV unless the results are not helped to make the specific medical diagnosis. Furthermore, in the last few years frequency domain technique spread out to find out the fluctuation patterns on the RR interval in time series, and it helps in clinical diagnosis and prognosis [3]. Most of the HRV measurement studies predict the risk of mortality (prognosis). But there are only a several studies which focused on diagnosis purpose [1].

The HRV data is extracted from electrocardiogram (ECG) records. The central spike of ECG record is QRS complex seen in ECG. Since R peaks in QRS complexes represent the ventricular contractions, the beats instants are taken at these points. The beat-to-beat intervals are calculated by measuring the time distance between R waves. Therefore, the term "beat-to-beat interval" known as an RR interval which is an unevenly sampled data [1]. Clayton et al.[4] described that FFT and AR methods can estimate the comparable low-frequency (LF) and high-frequency (HF) metrics for linearly sampled data (the power between 0.04 to 0.15 Hz and 0.15 to 0.4 Hz, respectively) [5]. To handle with unevenly sampled data issue RR interval data needed to be interpolated by the cubic spline and sampled at 4 Hz for Fast Fourier Transform (FFT) and Autoregressive Model (AR) because both techniques require evenly sampled data [3].

Besides, it is essential to remove the non-sinus (ectopic) beats from unevenly sampled RR tachogram because ectopic beats introduce further resampling problems and increase the variance of the interbeat intervals [5]. The ectopic beat should be removed from RR tachogram for HRV analysis because variability in the RR intervals may occur due to the ectopic beat. Usually, ectopic beat often occurs earlier and sometimes later than normal rhythms to happen. These unwanted short RR intervals make higher than standard frequency component. As a result, HRV measurement can be distorted from the actual measurement [2]. Birkett et al. [6] computed HRV spectra for congestive heart failure patients in two ways; firstly, ectopic beats were replaced by linear and cubic spline interpolation [7]. Secondly, ectopic beats were discarded. They concluded that high Frequency (HF) components were unaffected but low-frequency components (LF) was much higher using either interpolation method. In 1994 Lippman et al. [8] observed the linear, cubic, nonlinear prediction interpolation and the null case. After that, they summarized that ectopic correction should take into account for HRV analysis. Their finding was that linear, and cubic interpolation can interfere in power spectral estimations.

Furthermore, Lomb-Scargle (LS) method shows better performance to estimate the PSD [9, 10, 11] and it is analyzed by Moody et al.[12, 13]. Their result shows that PSD estimation is more accurate for typical RR tachograms in LS method compared to the FFT method. Because of it is possible to use without any resample and detrend RR data in LS method [10, 6].

In addition, according to the power spectrum density, spectrum entropy based on the Shannon entropy is defined in the frequency domain. Spectrum entropy is a measure of the unsteady signals and complexity of the system at frequency plane. Zheng et al. [14] presented the six wavelet entropies including the WEE, WTE, WSE, WTFE, WAE and WDE based on the wavelet theory and information entropy. Wavelet entropy measure used in EEG signal, mechanical fault diagnosis, and transmission line fault detection and identification. Besides, Isler Y et al. [1] described the HRV index uses which depend on the wavelet entropy measurement. They provided the three-step procedure to calculate the wavelet entropy and concluded that wavelet entropy calculation significantly improves the performance of the HRV analysis in the diagnosis and discrimination the congestive heart failure (CHF) patients from healthy patients. In this study, power Spectrum entropy measure is also included. Spectrum entropy is obtained based on the Shannon entropy definition. After that, this result is used to discriminate between different arrhythmias.

The aim of this study is to compare between different spectral estimation methods for missing data and exploring the suitable way when applied to HRV frequency domain analysis through Monte Carlo simulation using the R-R interval model. In this simulation, several spectral estimation techniques are used for PSD estimation. Besides, the long-term normal, AF, tachycardia and bradycardia patient's R-R interval tachograms are considered for the simulation of missing data for observing the missing data effect. The simulated gap generated based on the Poisson exponential distribution because actual missing data follow the Poisson exponential distribution method which obtained from ECG measurements carried out in research partner Indian hospital, they full filled all required ethical issue. Also, this study measures the spectral entropy for missing data to differentiate between arrhythmias.

The structure of the thesis is as follows. This thesis is divided into six chapters. Chapter 2 describes the Heart anatomy and generation of the ECG signal. Besides, it summarizes the related literature based on the different type of arrhythmias. Chapter 3 explains the used methods in this work in details. Firstly, it describes the method which we used to extract the RRI signal. After that, this chapter explains the spectral-based approach for estimating the PSD using different percentage missing data. Then, it describes the power spectrum entropy estimation based on PSD value. In chapter 4, we obtained experimental results from the collected database using different methods. Chapter 5 presents the validity of result and future work. Finally, we summarize our thesis work in chapter 6.

2. ARRHYTHMIAS AND BASICS OF ELECTROCARDIOGRAM

2.1. Basics of ECG

Electrocardiogram (ECG) is one kind of non-invasive implement for detecting cardiovascular diseases. ECG is used to assess the electrical and muscular functions of the heart. It can be measured from the body surface using the array of electrods [16].

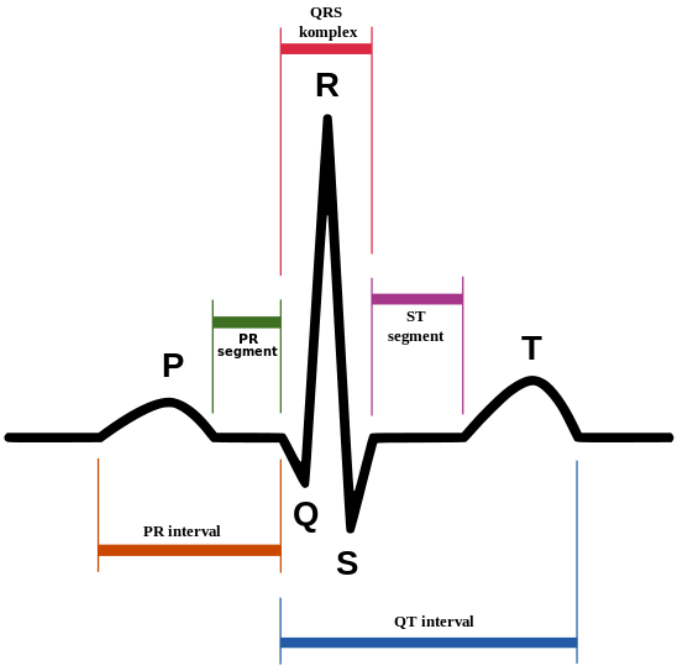

Figure 1. Represents a typical ECG waves [15].

The heart is consist of four chambers- right and left atrium and right and left ventricle. The right side of the heart accumulates the blood and pumps it's to lungs. After that, the left side of the heart receives blood and pumps it's to the body. Extracted ECG signal from the heart has three main waves like P, QRS complex, and T waves. Each wave is constructed due to specific reasons of heart activities [17]. P wave comes due to atrial depolarization of the heart. Then, the QRS complex is generated by ventricular depolarization of the heart chamber. In the same way, T wave shows the ventricular repolarization. Besides, U waves represent repolarization of the papillary muscles. It is small waves, and it does not present always in the ECG signal [18].

An electrocardiogram is the measurement of electrical activity of heart muscle. These can be obtained from the skin and the different angles like figure 2. Heart muscle collects and pumping the blood for all parts of the body. As a result of this, an action potential is generated by heart muscle which is known as electrical activity.

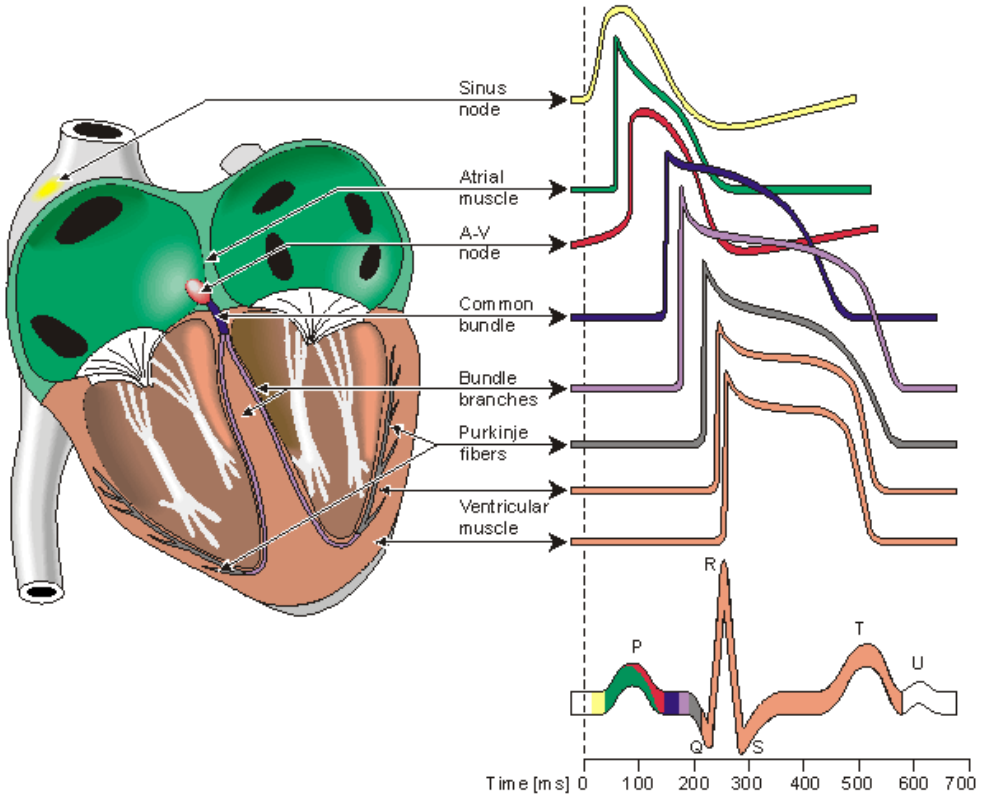

Figure 2. The heart anatomy with waveforms from different specific part of the heart [19] .

2.2. Arrhythmias

An arrhythmia is an irregular heartbeat. An arrhythmia is one kind of symptom of a wide variety of diseases, disorders which differentiate the normal and abnormal rhythm. Doctors can diagnosis by observing the ECG. For normal healthy subjects, the ECG looks at the periodic signal with depolarization followed by repolarization at the same intervals. Different ages people can be suffered from arrhythmias. An arrhythmia can occur due to smoking, electrolyte imbalance, alcoholism, drinking too much caffeine and excessive exercise. There are different types of arrhythmias. It can be mild to severe. Even it can be life threatening. Ventricular tachycardia and ventricular fibrillation cause for life-threatening arrhythmias. Paroxysmal atrial tachycardia, atrial flutter and second-degree heart block are also reasons for serious types of arrhythmias [20].

In here, we represent some common arrhythmias in details for our work. Accepted normal heart rate limits are 60 to 100 beats/min for normal sinus rhythm [21]. For normal sinus rhythm, the P-R interval is the first measurement. It is measured between the beginning of the upslope of the P wave and the beginning of the QRS wave. The P-R interval measurement should be from 120ms to 200ms. The R-R interval between the two cycles is constant for normal sinus rhythm [21]. As a result, rhythm must be regular.

Then, tachycardia refers to the heart rate that is too fast. Sinus tachycardia is normal which increase the heart rate, but the heart beats properly. Tachycardia occurs when the

heart produces the rapid electrical signal which is faster than normal while at rest. In this condition, the natural pacemaker of the heart, sinoatrial (SA) node produces faster electrical signal than usual [22]. Tachycardia frequently occurs in exercise, anxiety, fright and several emotional distress. It also can occur due to medical problems like anemia, increased thyroid activity, and severe bleeding. In this condition, the heart rate is fast, but heartbeats are properly. That is why R-R interval is constant and regular. On the other hand, the R-R interval is shorter than the normal sinus due to the fast heart rate. And P-R interval measurement should be from 120ms to 200ms [23].

After that, bradycardia is a slower heart rate than normal. It relies on the age and physical condition. Generally, the resting heart rate of fewer than 60 beats per minute (BPM) for adults is known as bradycardia. Causes of bradycardia are heart tissue damage and infection, imbalance of potassium and calcium in the blood, hypothermia, sleep, and some drugs. As like tachycardia patients, the heart beats are properly. Therefore R-R interval is constant and regular. But heart rate is slower then tachycardia and normal rhythm. As a result, the duration of the R-R peak is more prominent compared to the tachycardia and sinus rhythm [23].

Finally, Atrial fibrillation (AF) is a common cardiac arrhythmia which is used to describe an irregular and abnormal heart rate in the elderly. It is characterized by a disorganized electrical signal in atria. AF has serious health implications. It increases the risk of having a stroke [24]. Accepted normal heart rate is from 60 to 100 beats per minute for normal people. But this number usually goes over the 100 in AF patients. AF can occur due to the high blood pressure and overactive thyroid. AF can affect at any age. But it is rare in children.

3. SPECTRAL ANALYSIS WITH MISSING DATA

In this chapter, the applied methodologies are described as follows. Firstly, section 3.1 describes the missing data handling. After that, the signal preprocessing techniques which prepare the ECG signal and extract the information for further analysis. The next step, R-peak detection is accompanied with further consideration including ectopic beat detection and editing. Furthermore, Monte Carlo simulation has been revealed in section 3.3. Besides, this chapter examines the different methods that have been used in this thesis. Section 3.4 presents the spectral based approaches which have been used in this study for missing data analysis. Finally, the last section describes the spectrum entropy for different types of spectral signal.

3.1. Missing Data Handling

In our study, we used the real-time project data which collected by research partner India. We recorded the 19 patients data for analysis. In this project, data were received by the wireless device with blue tooth connection. After the investigation, we have observed that there are the missing data due to blue tooth packet loss in real-world. Then, we have found the missing data duration percentage and non-missing data duration percentage from each patient whole recorded data. Finally, we have drawn the histogram based on the missing and non-missing data distribution of real recorded data. And we observed that most of the patient missing data distribution follow the Poisson distribution. For this reason, In this study, we used the exponential Poisson distribution method to generate the simulated missing data using the PhysioNet database. For producing the missing data we use the exponential distribution for inter-event intervals as the number of events is Poisson distributed in the following way:
The Poisson distribution with probability mass function is:

$$P(X) = \frac{e^{-\mu}\mu^{x}}{x!} \tag{1}$$

where x = 0, 1, 2, . . . , and μ is the mean distribution and e is the exponential. Here, the variance of this distribution is equal to the mean distribution [25]. On the other hand , exponential distribution is a continuous distribution with probability density function,

$$f(t) = \lambda e^{-\lambda t} \tag{2}$$

where $t \geq 0$ and the parameter $\lambda > 0$. In this distribution, the mean and standard deviation are both equal to $\frac{1}{\lambda}$ [25].
So the cumulative exponential distribution is,

$$F(t) = \int_{0}^{\infty} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t} \tag{3}$$

3.2. Measuring HRV from ECG

Raw ECG signals are contaminated with different types of artifacts. The top-of-the-artifacts are power line interference, motion artifacts, baseline wandering, and instrumental noise. Before analyzing the ECG signal, the influence of those artifacts should be suppressed. Especially, for RRI signal construction, baseline wander correction is an important one because it alters the signal's amplitude directly. Besides, others artifacts also make the problem for further ECG signal analysis. That is why all of those artifacts should be suppressed for further frequency-domain processing.

3.2.1. Baseline Wandering Reduction

In biomedical electric recordings, baseline wander is a common phenomenon. Due to this, baseline drifts from its original position. Specifically, in our database, it occurs mainly due to movement artifacts. Besides, respiration might alter the ECG baseline. To reduce the baseline wander, we pass the ECG signal through a third- order Savitzky-Golay filter. Afterward, as it is illustrated in figure 3, the baseline eliminated ECG signal is produced after the constructed baseline ECG signal is subtracted from the original raw ECG signal.

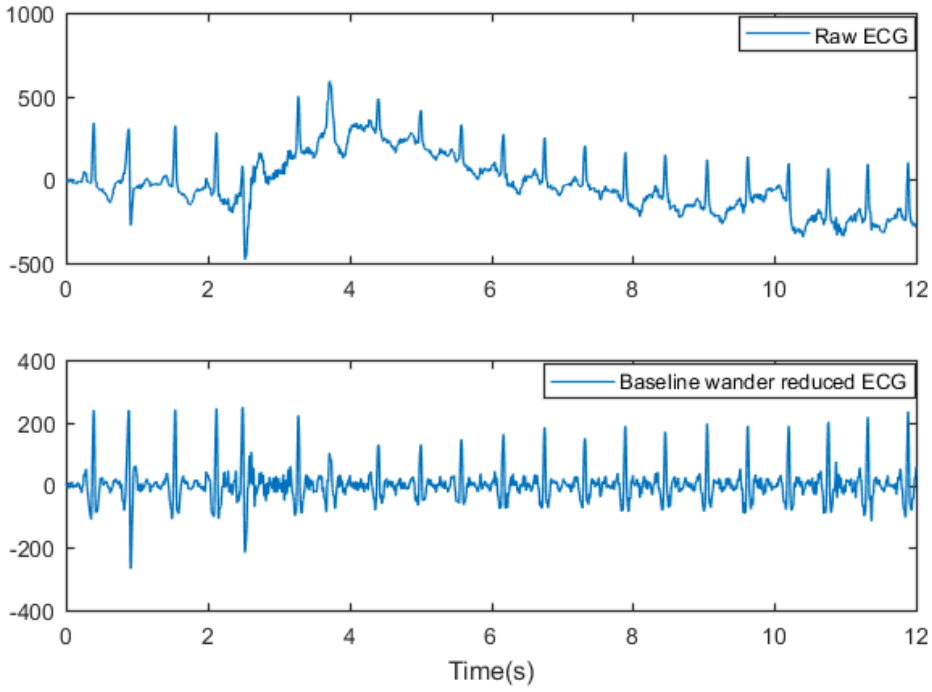

Figure 3. Baseline wander reduction sample figure using the Savitzky-Golay filter. The top sub-figure presents that how the amplitude alters of the signal due to the baseline. And the lower sub-figure describes the baseline reduced ECG signal.

Actually, based on the polynomial of a given degree, Savitzky-Golay filter does smoothing the contaminated data by executing the least square fitting of a frame of

data [26]. The peak preservation property of Savitzky-Golay has found attractive application in ECG data processing.

3.2.2. R-peak Detection

After removing the baseline wandering, the most important parts is the QRS complex detection in the ECG signal Analysis. QRS detection, especially R wave detection in heart signal is more accessible than other portions of ECG signal because structural form and amplitude of R wave are high. To detect R-peaks the ECG signal is inspected for local maximums with several constraints including having amplitude more than a predefined threshold, minimum peak prominence and minimum peak separation. Actually, R peaks detection provides the information of R-peaks amplitude and corresponding timestamps. That information is needed to construct the RRI signal and also need to analysis the power spectral using the fast Fourier transform, Autoregressive model and the Lomb-scargle periodogram method.

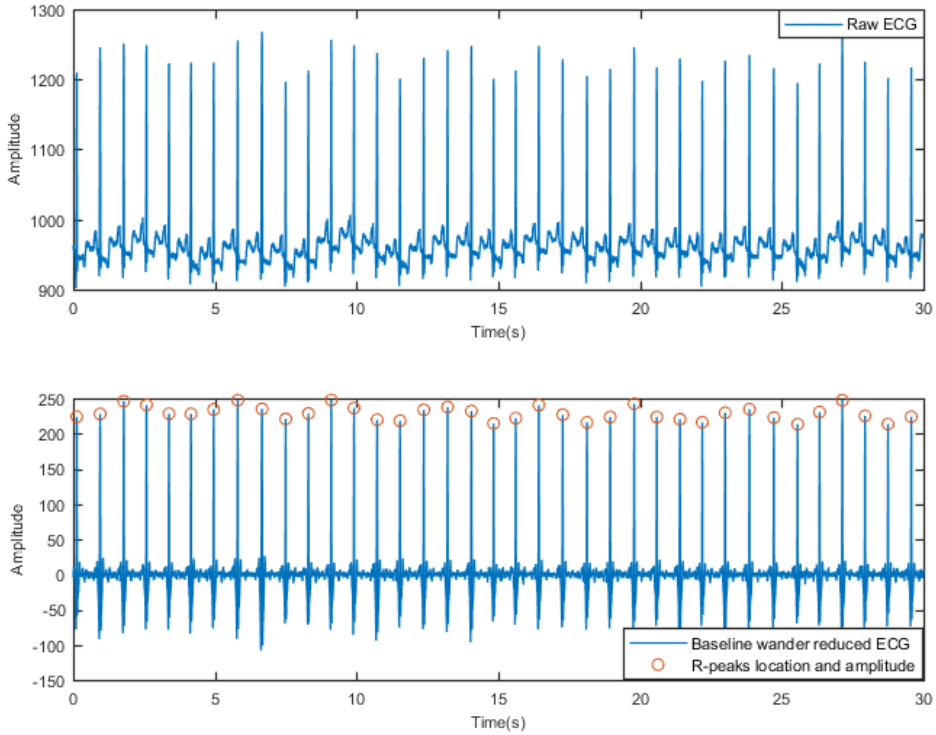

Figure 4. The procedure of R peak detection from baseline wander reduced ECG signal. The upper sub-figure represents the Raw ECG signal, and lower sub-figure shows baseline wander reduced ECG signal and detected R-peaks.

3.2.3. Ectopic Beat Detection and RRI (HRV) Signal Construction

For RRI signal construction, we need to detect the ectopic beats. There is a reasonable number of false, abnormal and ectopic beats in a long ECG signal. Physical activity inevitably increases the numbers of false beats. In this research, the heart rate variability (HRV) signal is anticipated to the normal R-R interval (RRI) modulated by sympathetic nervous (SN) system but can be disturbed by ectopic beats that are not generated from sympathetic nervous (SN) system. Mainly premature beats, tachycardia, and other arrhythmias are responsible for the ectopic beats. Besides, it can be by QRS wave misdetections. They distort the HRV signal with artificial frequency component. The ectopic beats could be detected from irregular RRI. There are some techniques to correct the ectopic confusion by deleting the beat or replacing the beat. In this research, we find the ectopic beat and replace it by NaN value. After that, for avoiding of substantial disturbance of power spectral density (PSD) estimation, we use the spline interpolation on the erroneous beats for FFT and AR method, but LS method does not require interpolation [27]. Figure 5 represents the constructed RRI signals using the cubic spline interpolation method. For RRI signal construction, we got R-peaks time annotations and the amplitude value of the signal. After that, we need to draw tachogram. For this, we used the time vector as R-peak timestamps, and the values are the difference between consecutive R-peak timestamps.

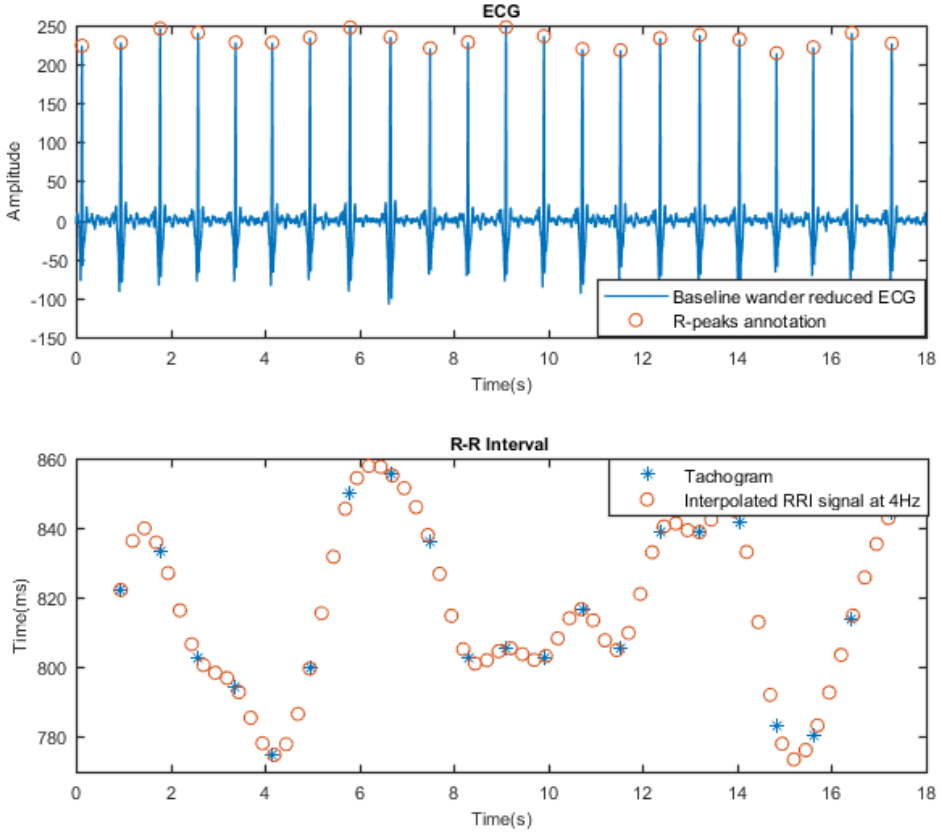

Figure 5. A sample set of the constructed signal. The upper sub-figure indicates the extracted ECG signal and red circle represents the R-peaks annotation. The lower sub-figure shows the constructed RRI signal and red circles represent the interpolated value using a cubic spline function.

3.3. Monte Carlo Simulations

In our study, Monte Carlo simulations are also used to determine the optimal spectral estimation method and observation of the missing data effect. For simulation, we used one thousand artificial R-R interval models. Based on the modified McSharry's models, these artificial models are generated randomly. The frequency parameters can be determined as the sum of each frequency range power. We obtained the true value of frequency parameters set from one thousand simulations can be represented as $[X_0^1, X_0^2, \cdots, X_0^N]$, where $N = 1000$ and X_0 is one of the true value of frequency parameters. The experimental results are calculated from the various spectral method with cubic spline interpolation method. These results can be presented as follows:

$$X = \begin{bmatrix} X_1^1 & X_1^2 \cdots & X_1^n \cdots & X_1^N \\ X_2^1 & X_2^2 \cdots & X_2^n \cdots & X_2^N \\ \vdots & \vdots \ddots & \vdots \ddots & \vdots \\ X_k^1 & X_k^2 \cdots & X_k^n \cdots & X_k^N \\ \vdots & \vdots \ddots & \vdots \ddots & \vdots \\ X_k^1 & X_k^2 \cdots & X_k^n \cdots & X_k^N \end{bmatrix}$$

where k= number of spectral methods * number of interpolation methods. So , in our study k = 3 * 1 +1 = 4.

The root-mean-squared relative error (RMSRE) is the method to determine the optimal method for calculating the HRV frequency parameters. Based on the 1000 artificial R-R interval models, absolute relative error's mean value of certain parameters between true values and experimental values is given as follows:

$$RMSRE[X_k] = \sqrt{\frac{1}{N}\sum_{n=1}^{N} RE[X_k^n]^2}\ ;$$

where,

$$RE[X_k^n] = \frac{X_k^n - X_0^n}{X_0^n} \cdot 100 \tag{4}$$

Here K represents $1, 2, \cdots, n$. If the RMSRE value is negligible, the k will be the optimal spectral analysis in this method [28].

3.4. Power Spectrum Estimation

Spectral-based methods aim to describe the distribution of power of a signal over frequency. The statistical properties of biosignals vary in time. That is why it is called non-stationary signals. To estimate the spectral density of constructed signals, we have used frequency analysis methods. For this purpose, among various method, we have used fast Fourier transform (FFT) method (Welch's Power Spectral Density), auto-regressive (AR) method (Burg's Power Spectral Density) and Lomb-Scargle (LS) method. These methods will be briefly introduced in this section.

3.4.1. Fast Fourier Transform

Power spectral density (PSD) can be calculated by the parametric and nonparametric method. Each of this method has different advantages. Fast Fourier Transform (FFT) is a non-parametric method for determining the PSD. It has a high processing speed and a simple algorithm. This method requires evenly spaced data for estimating the PSD. But R-R interval data is unevenly spaced sampled data. That is why these method has been suggested to interpolate the unevenly sampled R-R interval data. In this study, we interpolated the R-R interval series using cubic spline at sampled at 4 Hz [29].

Let, $x_a(t)$ is the sample version of x[n] with $-\infty < n < \infty$. If the sample rate is two times of maximum bandwidth frequency, Fourier transform expressed as:

$$X(F) = \sum_{n=-\infty}^{\infty} x[n]e^{-2j\pi fn} \tag{5}$$

Then, the energy density spectrum can be determined for the digitally sampled signal:

$$S_{xx}(F) = \mid \sum_{n=-\infty}^{\infty} x[n]e^{-2j\pi fn} \mid^2 \tag{6}$$

But for estimating the Fourier transform, Length of the x[n] should be finite. Then, spectral estimate of N points is,

$$\overline{x}[n] = x[n] \cdot w[n], 0 < n < N-1; \tag{7}$$

where w[n] is a rectangular window which multiplies with finite length signal x[n]. Now frequency relationship can be represented from equation :

$$\overline{X}(f) = X(f) * W(f)$$

These $\overline{X}(f)$ indicates the convolution of their Fourier Transforms which equal to two-time sequence's multiplication. Here, if the $X(f)$ spectrum is broader than the spectrum of $W(f)$, $\overline{X}(f)$ will be smooth. But leakage problem will remain for the convolution of the side lobes. On the other hand, when the $X(f)$ spectrum is narrower than the $W(f)$ spectrum, it brings two critical issues, one is leakage problem, and another one is spectrum resolution [30]. After that, for estimation of power spectrum density of the random signal is done by periodogram calculation, which can be expressed:

$$P_{xx} = \lim_{n\to\infty} E[S_{xx}] \tag{8}$$

$$P_{xx}(F) = \mid \sum_{n=0}^{N-1} x[n]e^{-2j\pi fn} \mid^2 \tag{9}$$

Using this equation to estimate the actual power spectrum is not a consistent measure because it has leakage problems and bias. So it can spoil the spectrum. As an alternative to removing those problems, Welch's method comes forward [29]. Based on this, we used Welch's method to estimate the PSD in our study.

3.4.2. Welch's Power Spectral Density

Welch's method is used for estimating the power of a signal at different frequencies. This method involves sectioning the signal, taking modified periodograms of these sections, and averaging these modified periodograms [31]. Based on the periodogram spectrum estimation, the method converts a signal from the time domain to the frequency domain. Welch's method is also called the averaging method and weighted overlapped segment averaging (WOSA) method [32]. In our study, we used the Hanning window for each segment. Its length is equal to the sampling frequency of the

signal. And we consider the 50 percent overlap between each segment. According to these parameters which have used in our study, we can write:

Let consider $X(j)$, j= $0, \cdots, N-1$ be a sample from a stationary, second-order stochastic sequence. And consider $P(f)$ is the spectral density of the $X(j)$. Now we partition the data sequence into segments, possibly overlapping length L with the starting point of these segments D units apart,$X(j)$, j=$0, \cdots, L-1$ indicates the first segment.

$$X_1(j) = X(j)$$

Where, j = $0, \cdots, L-1$. In the same way,

$$X_2(j) = X(j+D)$$

Here, j = $0, \cdots, L-1$. And finally,

$$X_k(j) = X(j+(k-1)D)$$

where, j= $0, \cdots, L-1$. Now for the k segments;

$$X_1(j), \cdots, X_k(j)$$

and they cover the entire record like,

$$(K-1)D + L = N$$

Now we can calculate a modified periodogram for each segment of length L.
For this, we select the data window $W(j)$, j=$0, \cdots, L-1$, and form the sequences, $X_1(j)W(j), \cdots, X_k(j)W(j)$.
Then, we can take finite Fourier transforms $A_1(n), \cdots, A_k(n)$ of these sequences. So,

$$A_k(n) = \frac{1}{L} \sum_{j=0}^{L-1} X_k(j) W(j) e^{-2kijn/L} \tag{10}$$

and i = $(-1)^{1/2}$, In the final stage, we get the K modified periodograms

$$I_k(f_n) = \frac{L}{U} \mid A_k(n) \mid^2 \tag{11}$$

where, f_n = n/L and n= $0, \cdots, L/2$ and,

$$U = \tfrac{1}{L} \textstyle\sum_{j=0}^{L-1} W^2(j)$$

The spectral estimate is the average of this periodograms,

$$P(f_n) = \frac{1}{K} \sum_{k=1}^{K} I_k(f_n) \tag{12}$$

Hence, $P(f_n)$ is the spectral estimator with a resultant spectral window. The window area is unity and width is of an order of $\frac{1}{L}$ [31].

3.4.3. Autoregressive Model

The spectral analysis describes the power of signal over frequency. A time-frequency representation used in our thesis is Brug's-based Autoregressive (AR) method. AR method is the parametric method. In the parametric method, there is an all-pole linear filter which discusses the signal's spectral components. If the PSD is large at certain frequencies, The AR method can describe in a good way. Besides, when the signal data is relatively short, The AR method generate the good result compared to the non-parametric methods. Moreover, AR method generates smoother PSD than non-parametric methods. In our study, we used the 11th order with the sampling frequency 4Hz. Now, we can represent our constructed signal in time-domain by a p-order AR model if the discrete time series x(t):

$$x(t) = \sum_{i=1}^{p} a_i x(t-i) + e(t) \tag{13}$$

where a_i are the AR model coefficient and e(t) indicates the estimation error. The model can be express as a z-transform:

$$X(z) = X(z) \sum_{i=1}^{p} a_i z^{-i} + E(z) \longrightarrow X(z) = H(z)E(z) \tag{14}$$

The PSD of z-domain can be stated as

$$P_x(z) = \mid H(z) \mid^2 P_e(z) \tag{15}$$

By putting of $z = e^{2\pi j f \Delta t}$, the power spectral density estimation can be stated as:

$$P_x(f) = \frac{\sigma^2 \Delta t}{\mid 1 - \sum_{i=1}^{p} a_i e^{(-2\pi j f \Delta t)} \mid} \tag{16}$$

where $e(t)$ is white noise with zero mean and variance of σ^2.
For finding the parameters from equation 16, we can use different algorithms. In our study, we used Burg's based method for estimation of the parameters [17, 33].

3.4.4. Burg's Power Spectral Density

Burg's method is developed for estimating the AR parameters that are based on forward and backward prediction errors. In our study, Burg's method used for AR spectral estimation that minimizes the forward and backward prediction error by constraining the AR parameters according to the Levinson-Durbin recursion [34]. Levinson-Durbin recursion predicts the linearity of an all-pole IIR filter with deterministic autocorrelation sequence. So it estimates the AR parameters, $A(Z)$:

$$H(z) = \frac{1}{A(Z)} \tag{17}$$

$$H(z) = \frac{1}{1 + a(2)z^{-1} + a(n+1)z^{-n}} \tag{18}$$

where $H(Z)$ is the predicted transfer function of the linear system. $A(Z)$ is the nth order polynomial and the filter coefficient a= $[1, a(2), ,,,, a(n+1)]$.

In our study, model order 11th is used. The model order indicates the number of poles of the AR model. It is essential to estimate the spectral information, and we can predict it from the input data. The burg's method shows the better result for estimation of short records data and better resolution of closely spaced sinusoids with low noise. Neophytos.N [35]. described that burg's method shows the small divergence of the PSD estimation from the true value. On the other hand, burg's method accuracy is decreasing for high model orders, more extended data, and higher signal-to-noise ratios. Finally, this method usually behaves somewhere between the LS method and the Yule– Walker method [35].

3.4.5. Lomb-Scargle Periodogram Method

The Lomb-Scargle method is used for performing the spectral analysis on missing sample data. The Lomb-scargle algorithm is well detecting and characterizing periodicity in missing sampled time-series. In our study, we used the lomb-scargle algorithm for calculating the PSD in missing sampled time-series. For classical methods, we can not estimate PSD directly for missing sampled data without resampling. This resampling can make artifacts in the estimated spectrum. On the other hand, for the Lomb-scargle method, we can calculate the spectrum directly for missing sampled data without any resampling or interpolation. And we can avoid spectrum distortion in the Lomb-scargle method [36]. In the Lomb method, PSD estimates based on the minimization of the squared differences between the signal and the basis function of the transform.
Let's consider the, $x(t)$ is the continuous signal and $b_i(t)$ is the basis function of the transform. The coefficients $c(i)$ that represent $z(t)$ in the transform domain. This minimize the squared error $e(c)$ defined as: $e(c)$ = $\int_{-a}^{b}(x(t) - c(i)b_i(t)^2)dt$.
Lomb develops model to estimate the Fourier spectra for missing samples at t_n instants by adjusting the following model, if the signal x(t) is accessible only at missing samples.The model $x(t_n) + E_n$ = $cb_i(t_n)$, in such a way that the variance E_n can be minimized for value c. So,

$$c(i) = \frac{1}{K}\sum_{n=1}^{N} x(t_n)b_i(t_n) \tag{19}$$

$$\text{where, } K = \sum_{n=1}^{N} b_i^2(t_n)$$

This equation indicates the generalized Lomb method to estimate the transform of missing samples data.
From this equation, we can write the signal power at index i of the transformation:

$$P_x(i) = Kc^2(i) = \bar{c}^2(i) \tag{20}$$

$$\text{where } \bar{c}(i) = c(i)\sqrt{K}$$

Now we can consider $K = N$ for only the Fourier Transform [37]. After that, It can be expressed as:

$$P_x(i) = p_x(f) = \bar{c}^2(i) \tag{21}$$

3.5. Spectral Entropy

Longtime sustainable cardiac arrhythmias which may be severe or fatal damage to the heart. As a result, normal rhythmicity can be changed of the heart. Cardiac arrhythmias like atrial fibrillation, tachycardia, and bradycardia always lead to sudden cardiac death. It is essential to take immediate diagnosis for the patients if they suffer any kind of cardiac arrhythmias [38]. Electrocardiogram signal provides information about the cardiovascular system. In this study, we used the short term ECG signals to differentiate between arrhythmias based on the power spectrum entropy feature. Previously, there are several studies which described that power spectrum entropy (PSE) is an excellent way to extract the feature from Electroencephalogram (EEG) signal [39, 40, 41, 42, 43]. Kamath C. [38] described that they found the potential discrimination between healthy and VT/VF subjects using the spectral entropy features. They used the FFT method for estimating the spectrum and found the prognostic value for cardiac analysis. Besides, Zhou P.et al. [44] was observed the variation of power spectral entropy in different sleep stages by extracting the features from ECG signal, and they also described the regularity of nocturnal frontal lobe epilepsy. Furthermore, driving distraction was analyzed by an entropy analysis using ECG signal in Yu L.et al [45]. Their research explored that power spectrum entropy in driving with distraction was more substantial compared to without distraction. After that, spectral entropy also used for speech recognition automatically [46]. They showed that spectral entropy of spectrum could be used as discriminate of voicing/unvoicing. Also, Zheng-you H.et al [47] described the different types of wavelet entropy definition and its application to identify and detect the transmission line fault. Besides, there is a study which used HRV indices with wavelet entropy measures to discriminate the CHF patients from healthy normals.

The entropy of a signal is the measure of the uncertainty of a random variable. Actually, entropy measures the flatness of the power spectrum. In our study, we have used three methods like FFT, AR, and Lomb-Scargle method for estimating the PSD. Based on those spectra, entropy is used to measure the 'peakiness' of distribution after applying the probability mass function (PMF) [48]. In this research, we want to explore peaks capturing property of the entropy using the peaks of a spectrum. The entropy value will be low when a PMF with sharp peaks. On the other hand, a PMF with flat distribution will have the high entropy. When we want to use entropy to capture the peaks of a spectrum, the problem is that spectrum is not a PMF (the area under the spectrum does not sum up to 1). Spectrum entropy is measured in the frequency domain. If $Z(w)$ is the DTF of signal $x(n)$, the power spectrum is : $X(w) = \frac{1}{2\pi n} \mid Z(w) \mid^2$. If $x = (x_1, \cdots, x_N)$ is the partition of original signal, the i-th power spectrum possessed in whole is,

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i} \tag{22}$$

where X_i represents the energy of the i^{th} frequency component, N is the number of points in the spectrum.

This equation converts the spectrum into PMF function in a ways divided individual frequency components of the spectrum by sum of all the components. It ensured that

the normalized spectra is in PMF for the purpose of estimating entropy [49, 50]. Now we can calculate entropy for each frame from x by:

$$Entropy(H) = -\sum_{i=1}^{N} X_i \cdot log_2 x_i \tag{23}$$

4. EXPERIMENTAL RESULTS

In this chapter, we present the experimental result which obtained using different spectral based methods. Initially, we describe the structure of our data which used in this experiment. Then, the result of spectral based methods which used for missing data evaluation are described. After that, the result of each method is stated. Finally. We present the spectral-based entropy result which differentiate between different arrhythmias.

4.1. Real Missing Data

In our study, we used real data which collected by research partner from Indian hospital for investigating the effects of the real missing R-R interval data on HRV analysis and finding the optimal spectral method with missing data. Figure 6 provides the missing data and non-missing data characteristics of real data sets which estimate using the Poisson distribution parameters.

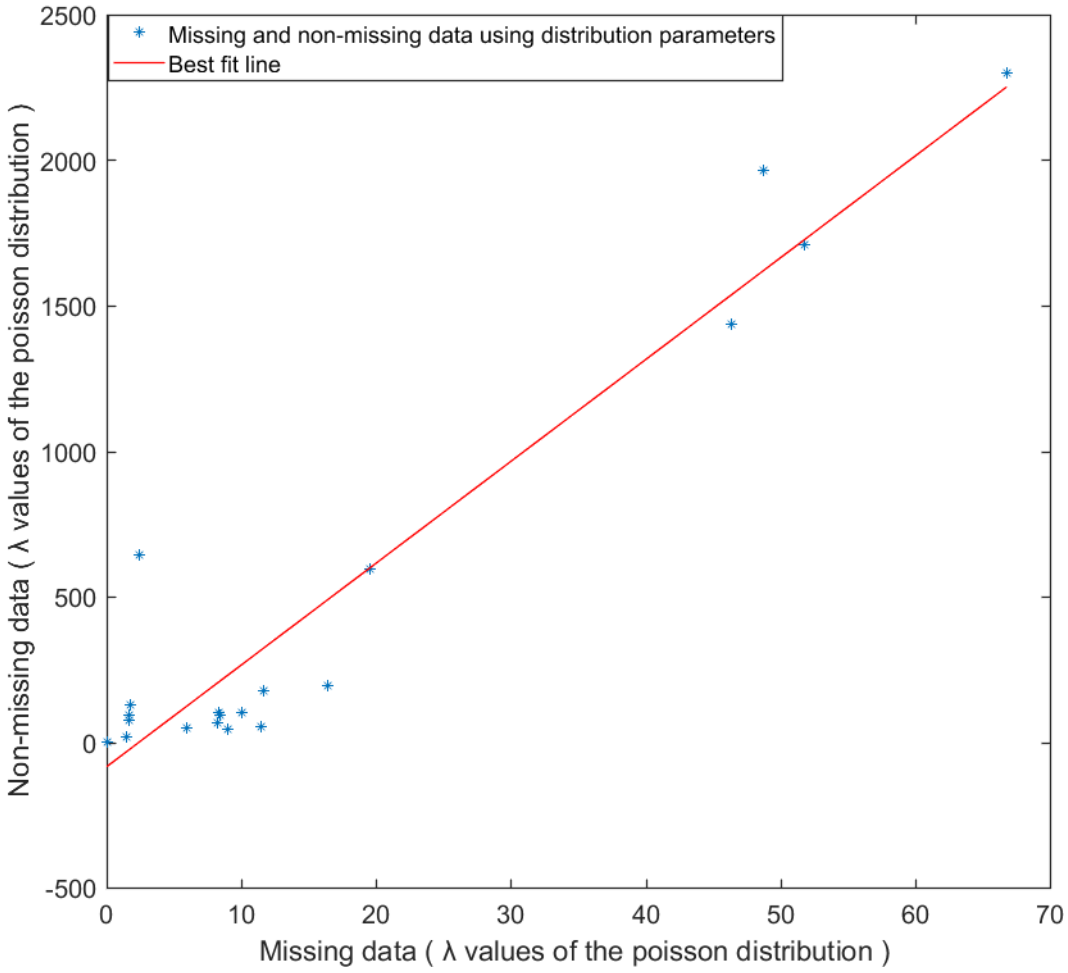

Figure 6. Missing data Vs. Non-missing data characteristics of real data sets estimating using the Poisson distribution parameters.

In the real collected data which has ethical permission from the authority, there are 19 patients about 6 hours recorded ECG with motion artifact and missing parts which known as NaN in data sets. All subjects were cardiac disease patients. But most of the patients was arrested by atrial fibrillation, tachycardia and bradycardia heart diseases. R-R interval data were extracted by using the R-peak detection algorithm from the original ECG data. But motion artifacts are dominant in real ECG; hence, it makes the complexity to detect R-peaks correctly from real data sets. For this reason, we calculated the non-missing data part with missing data part from real data sets.

After that, we calculated the consecutive missing run length from those 19 subjects. The missing run length distribution sample in the data sets is showed in figure 7. Besides, figure 7 presents the simulated gap length distribution based on the real missing data length characteristic. This missing data length in real data sets follows the exponential poison distribution.

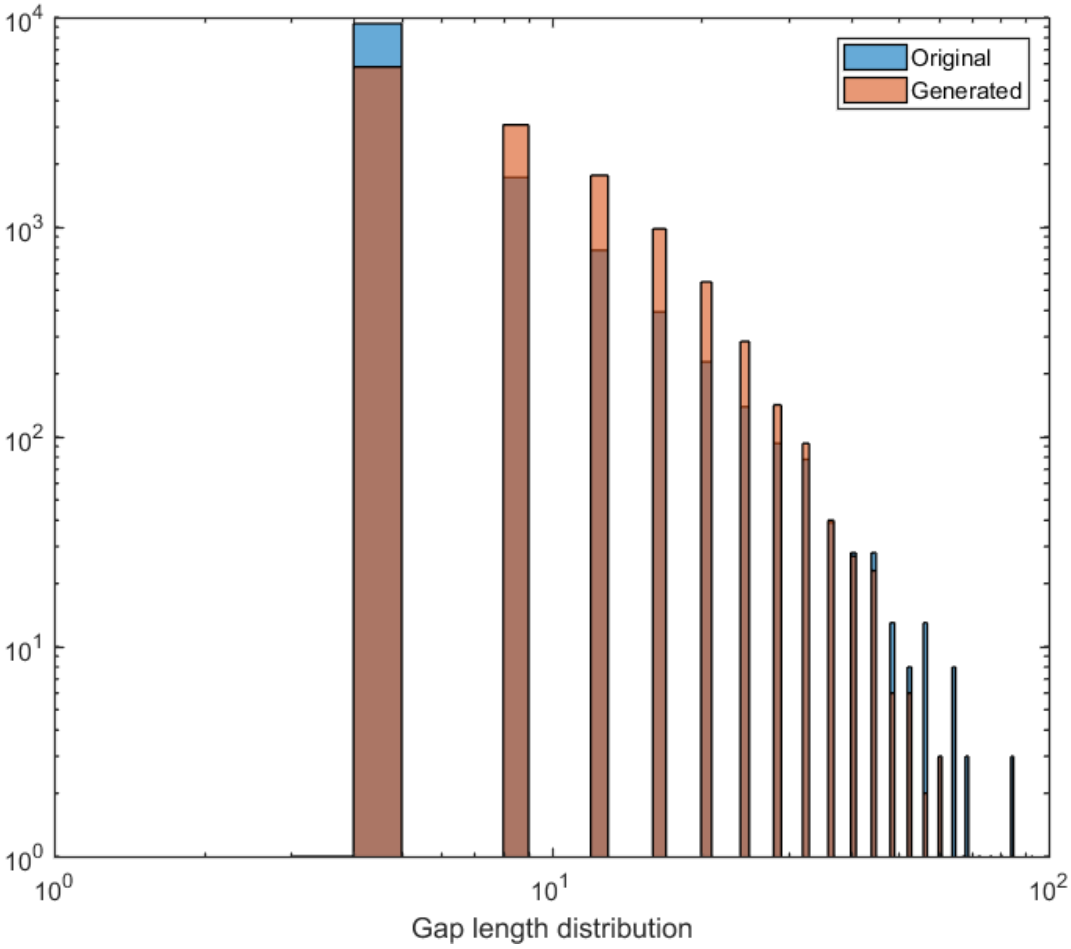

Figure 7. Log-normal consecutive gap length distribution of real sample data and simulated gap length distribution.

4.2. Simulation of the Missing R-R Interval Data

For this study, the long-term R-R tachograms data were collected from the MIT-BIH arrhythmia database (https://www.physionet.org/physiobank/database/mitdb/). There are 48 half-hour excerpts of two-channel ambulatory ECG recordings in the MIT-BIH Arrhythmia Database. After studying the BIH Arrhythmia Laboratory, 47 subjects obtained from 48 subjects. Twenty-three subjects from 47 were selected at random from a set of 4000 24-hour ambulatory ECG recordings from a mixed population. And the rest of the 25 subjects were chosen from the same set but clinically significant arrhythmias. Besides, the ECGs were digitized at 360 samples per second in every channel with 11-bit resolution and 10mV range. Moreover, two or more cardiologist annotated each record separately. And their disagreements were resolved using the computer-readable reference annotations for each beat. In this study, we used selectively extracted 5 minutes of ECG data including normal sinus rhythm, AF, tachycardia, and bradycardia patients for HRV analysis using previously mention three spectral based methods in chapter 3.

For the simulation, the data were removed by randomly from the R-R interval data. And randomly missing data duration remained between about 0 to 6 percentage based on the real missing data. Applying 1000 Monte Carlo runs using the MATLAB program, the HRV spectral parameters were estimated. In this study, the relative errors (REs) was used to observe the effect of missing data on HRV parameters in those three methods. The relative errors (REs) compared the missing data parameters with the original data parameters which have no missing data. The HRV parameters from the missing data is $[X_1, X_2, \cdots, X_n]$ where n= 1000 and the $X_original$ represents the parameter value from the non-missing data. Then, relative errors are calculated as, $RE s(\%) = \frac{X_{origin} - X_k}{X_{origin}} \cdot 100$ where $K = 1, 2, \cdots, n$.

4.3. Quantitative Results

In this section, we inspect the performance of different described methods for missing R-R interval data. Here, we explore the acquired result with the spectral-based approaches. We have used RRI signal in spectral-based approaches that RRI signal is collected from normal, AF, tachycardia and bradycardia patient signal. We illustrate the power spectral entropy for differentiating the arrhythmias in the latter part of this section.

4.3.1. Experimental Results for Normal Patient

In the beginning, figure 9 to figure 11 provides the experimental results for normal patient. In 9 and 8 figures, the blue line shows the power spectrum for FFT, AR, and LS based method respectively in sub-figure 1, 2 and 3 for original data without missing data duration as the reference values. In the same figures, the orange line represents the power spectrum estimation for those three methods for the percentage of missing data. Actually, in this research, we generated 1000 times artificial gaps in the R-R in-

terval tachograms for spectral-based approach analysis. This two spectrum from without missing data and with missing data is used to identify the suitable spectral-based method and the investigate the effect of missing data. Then, it is clear that figure 9 and 8 represent the PSD mismatch between non-missing data and missing data for FFT, AR, and LS based method. From those figures, it has been shown that the maximum peak point and PSD amplitude value are shifted for missing data based on the missing data percentage.

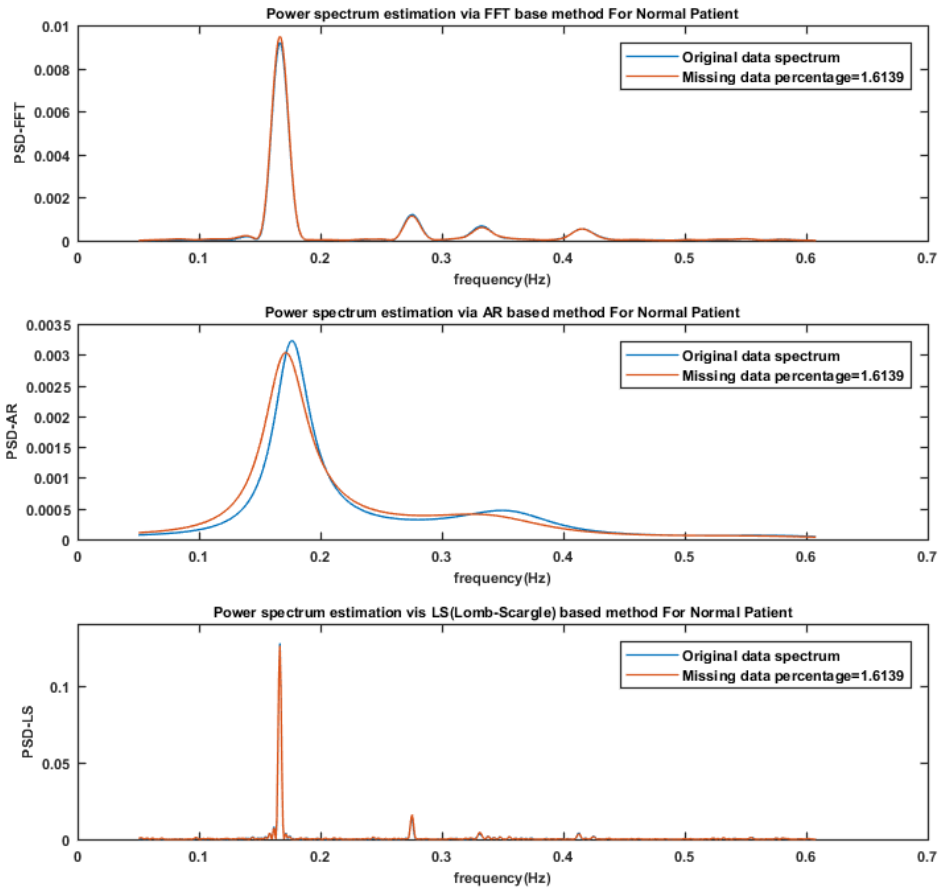

Figure 8. PSD mismatch between non-missing data spectrum and missing data spectrum for normal patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 1.61 percentage missing data.

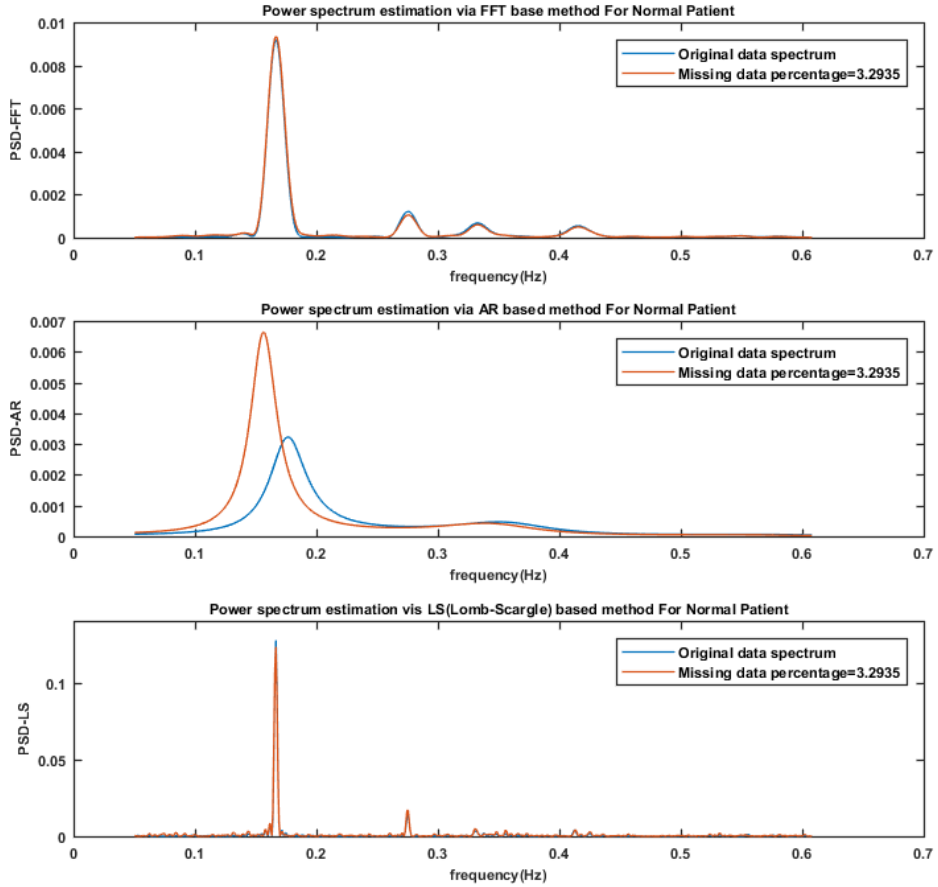

Figure 9. PSD mismatch between non-missing data spectrum and missing data spectrum for normal patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 3.29 percentage missing data.

Furthermore, figure 10 provides information about data distribution for the normal patient based on the duration of missing data via FFT, AR, and LS-based methods corresponding sub-figure 1, 2 and 3. This data distribution is obtained from 1000 Monte Carlo simulation results for each parameter, and the error power percentage is calculated based on each element difference using the equation 4. From this figure, it is visualized that all three methods of experimental result, the LS method is better with less error compared to other methods based on the effect of missing data percentage. Because the LS-based method shows the more compact and no more outliers as the percentage of missing data increased, whereas FFT and AR methods show the more outliers.

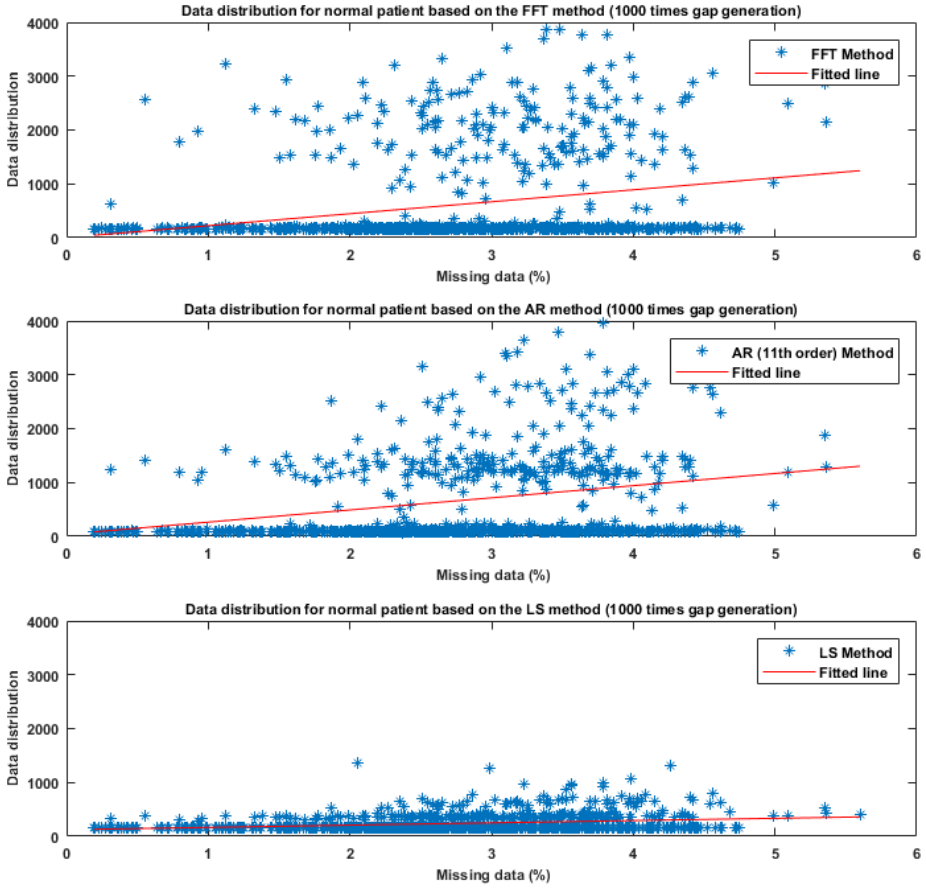

Figure 10. This plot shows the data distribution of error power based on the percentage of missing data. Besides, it describes distribution of random points comparison generated by Monte Carlo simulation for normal patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line.

After that, figure 11 shows the error power percentage between non-missing data spectrum and missing data spectrum based on the duration of missing data via FFT, AR, LS-based method respectively in sub-figure 1, 2 and 3 using box plot. The missing data duration axis expresses as a bin in which each bin contains 200 samples among 1000 Monte Carlo simulation samples. In this figure, the error power percentage for each HRV parameter was altered by different spectral estimation methods. In the sub-figure 1 of figure 11, it has been clear that median error remains almost same up to 3.06 percentage of missing data but median error increases gradually when missing data percentage goes above 3.06 for FFT method.

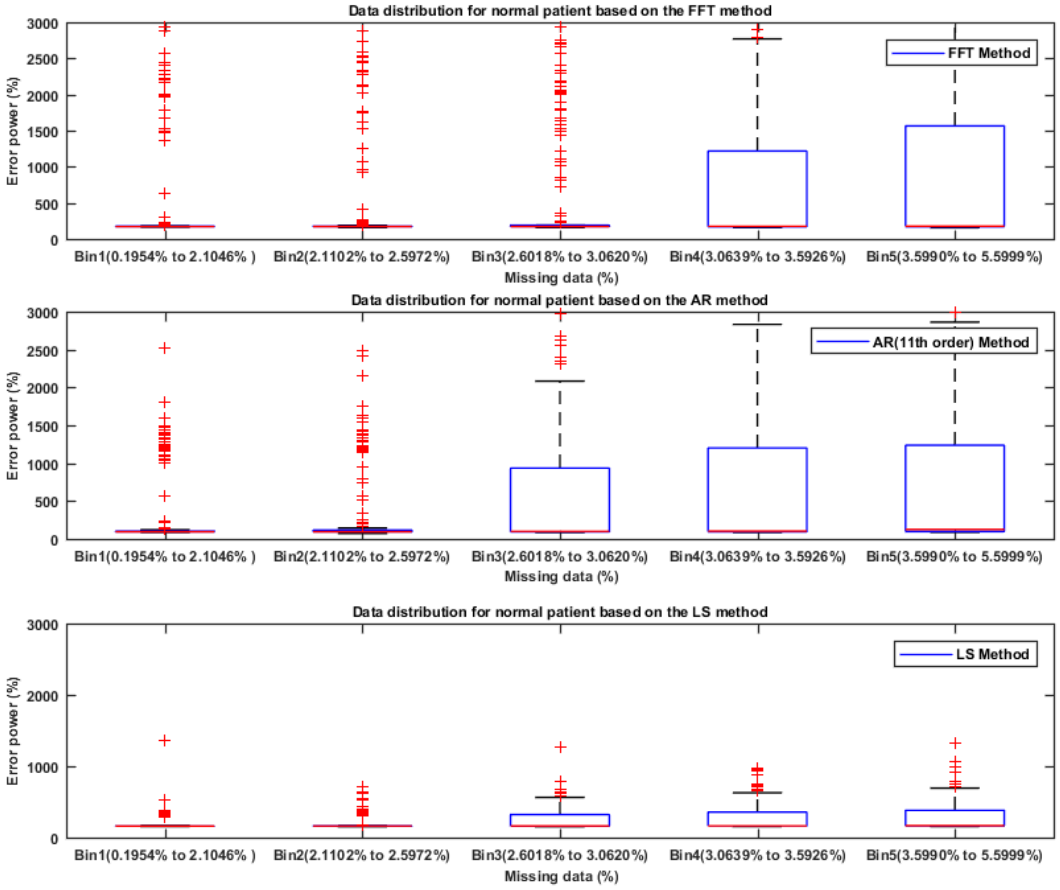

Figure 11. Boxplot of the subfigure 1, 2 and 3 estimated the error power percentage based on the missing data percentage for the normal patient via FFT, AR, and LS method respectively using 1000 times Monte Carlo simulated data. First quartile, median, and third quartile values are portrayed as the bottom, middle and top horizontal line of the boxes. Whiskers are used to represent the most extreme values. And outliers were displayed as crosses. The error power percentage was calculated by the equation 4.

The same trend is followed in sub-figure 2 for AR method. In this case, the median error rises when the missing data percentage cross the bin2, amounting 2.60 percentage missing data. On the contrary, in sub-figure 3 for LS method, as the percentage of missing data increases, the percentage of error power remains almost the same. Overall, in the LS method, there is less effect for missing data percentage compared to the FFT and AR method and number of outliers are fewer also. From this figure, it is clearly noticed that the error power percentage shows less amount between three estimates with LS for normal patient data in term of the missing data percentage.

4.3.2. Experimental Results for AF Patient

Now experimental results are described from figure 12 to 15 for AF patient. Actually, in this research, we used the same technique for AF patient data. Figure 12 and 13 estimate the power spectrum for above mention three methods respectively non-missing data without missing data as the reference values with the blue line and the orange line indicates the percentage of missing data spectrum.

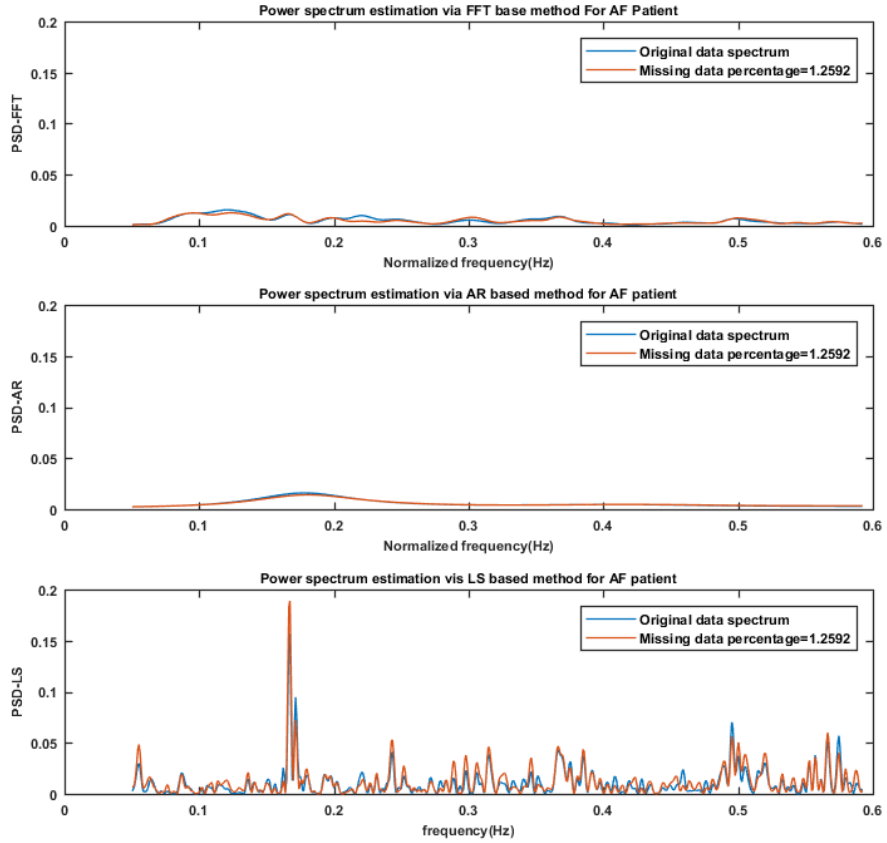

Figure 12. PSD mismatch between non-missing data spectrum and missing data spectrum for AF patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 1.26 percentage missing data.

After that, these two figures show the PSD mismatch between non-missing data and missing data for FFT, AR and LS based method for better understanding the mismatch error. From these figures, it has been shown that the maximum peak point and PSD amplitude value are shifted for missing data based on the missing data percentage. And the power spectrum is much flatter compared to the normal patient's power spectrum.

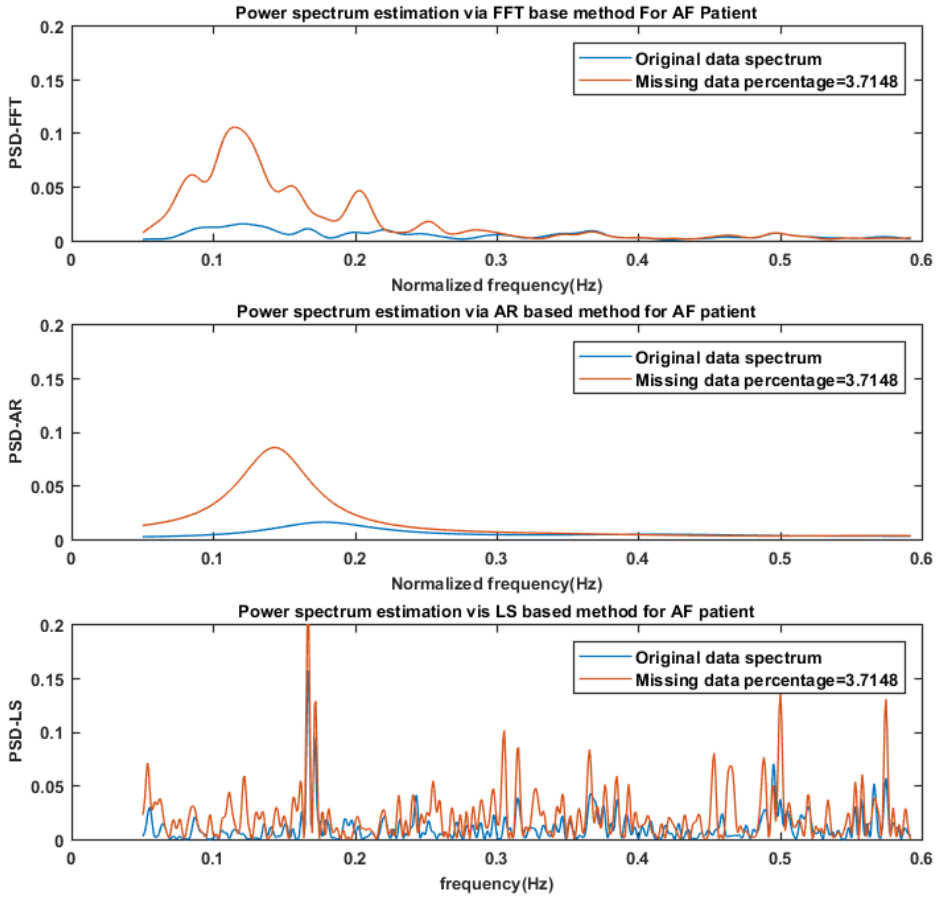

Figure 13. PSD mismatch between non-missing data spectrum and missing data spectrum for AF patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 3.71 percentage missing data.

Then, figure 14 illustrates that data distribution of error power based on the missing data percentage using the FFT, AR and LS method. From this figure, it is cleared that as the percentage of missing data increase, the data distribution of error power is more compact and remain almost the same line for LS method which illustrated in sub-figure 3. On the other hand, the data distribution of error power rises up when the missing data percentage increase. However, this figure indicates that the effect of missing data was much less in LS method compared to the others two methods for AF patient.

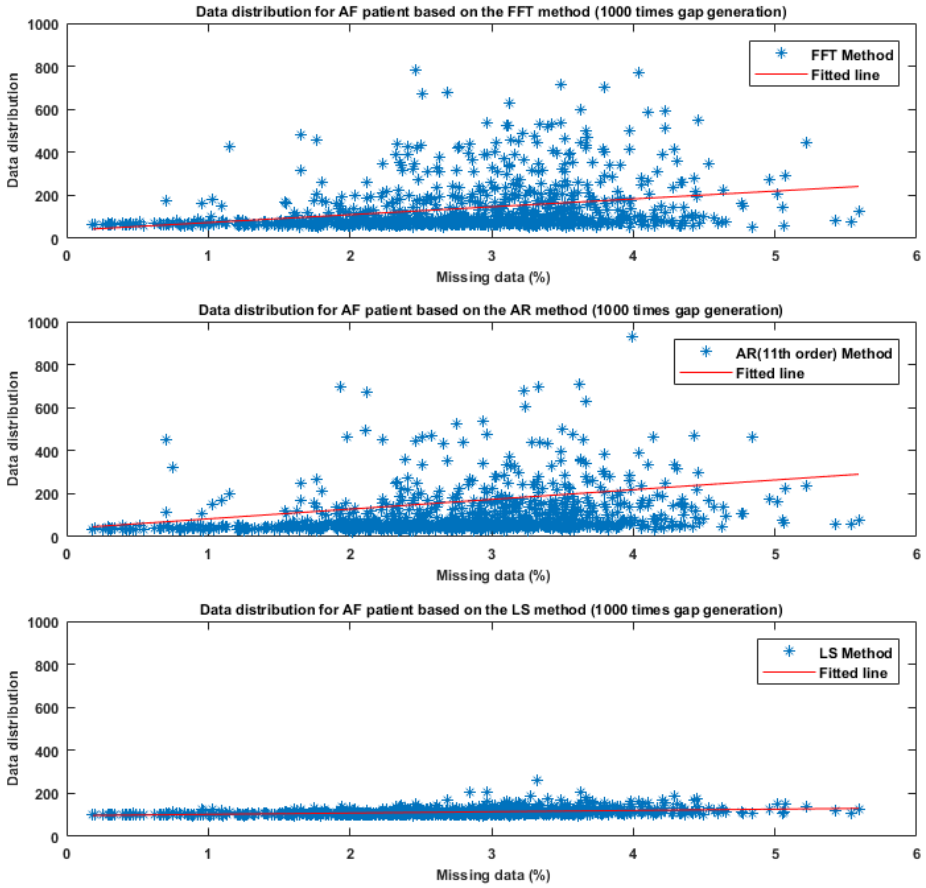

Figure 14. This plot shows the data distribution of error power based on the percentage of missing data. Besides, it describes distribution of random points comparison generated by Monte Carlo simulation for AF patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line.

Finally, figure 15 represents the error power percentage between the non-missing data spectrum and missing data spectrum for AF patient via previously described three methods based on the duration of missing data. In sub-figure 1 and 2, the median error is almost the same till 2.60 percentage of missing data for both cases namely FFT and AR method. After that, in both instances median error increases slowly when the missing data percentage goes above 2.60 percentage in sub-figure 1 and 2 respectively, whereas median error remains almost constant for LS method in term of the missing data percentage in sub-figure 3. Besides, the number of outliers is less in LS method compared to the other two methods namely FFT and AR method.

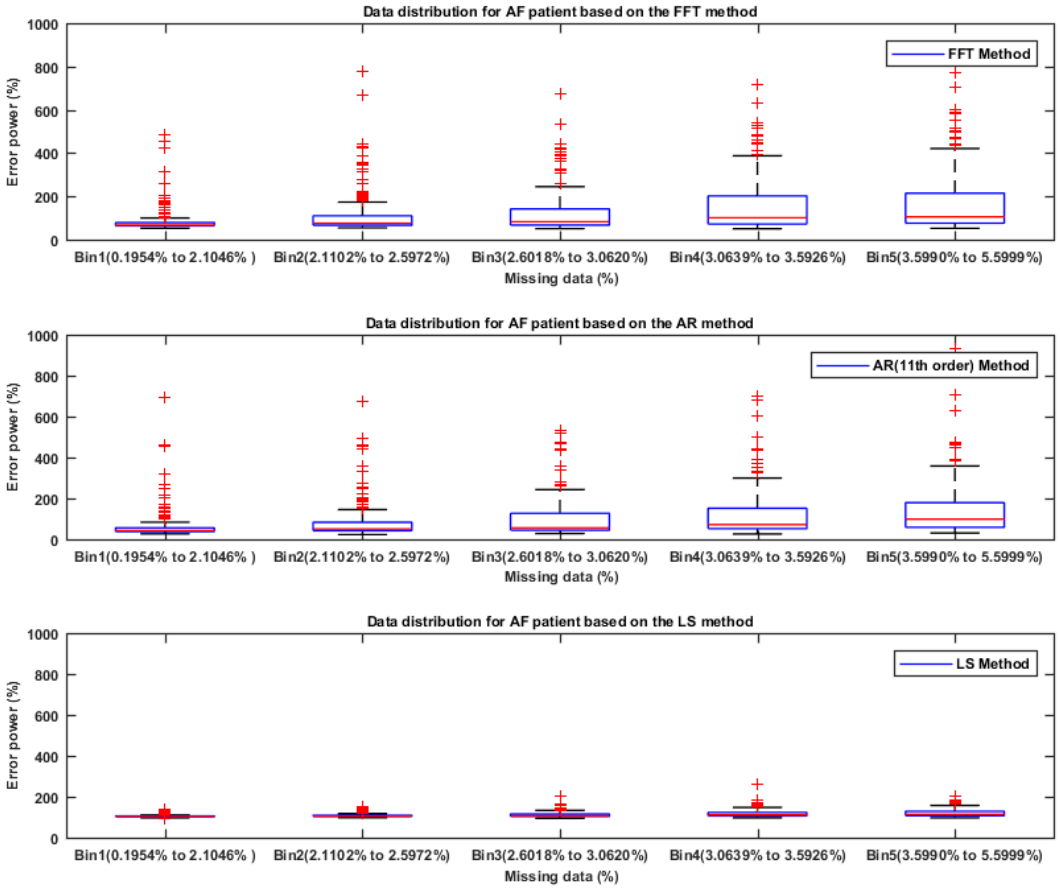

Figure 15. Boxplot of the subfigure 1, 2 and 3 estimated the error power percentage based on the missing data percentage for the AF patient via FFT, AR, and LS method respectively using 1000 times Monte Carlo simulated data. First quartile, median, and third quartile values are portrayed as the bottom, middle and top horizontal line of the boxes. Whiskers are used to represent the most extreme values. And outliers were displayed as crosses. The error power percentage was calculated by the equation 4.

4.3.3. Experimental Results for Tachycardia Patient

For tachycardia patient, experimental results are portrayed from figure 16 to 19. In figure 16 and 17, blue line and orange line represent the power spectrum respectively non- missing data as the reference values and percentage of missing data. Then, these two figures visualize the PSD mismatch between original data and missing data for FFT, AR and LS based method. It has been shown from those figures that the power spectrum is more flatter then the normal patient and the maximum peak point and PSD amplitude value are shifted for missing data corresponding to the missing data percentage.

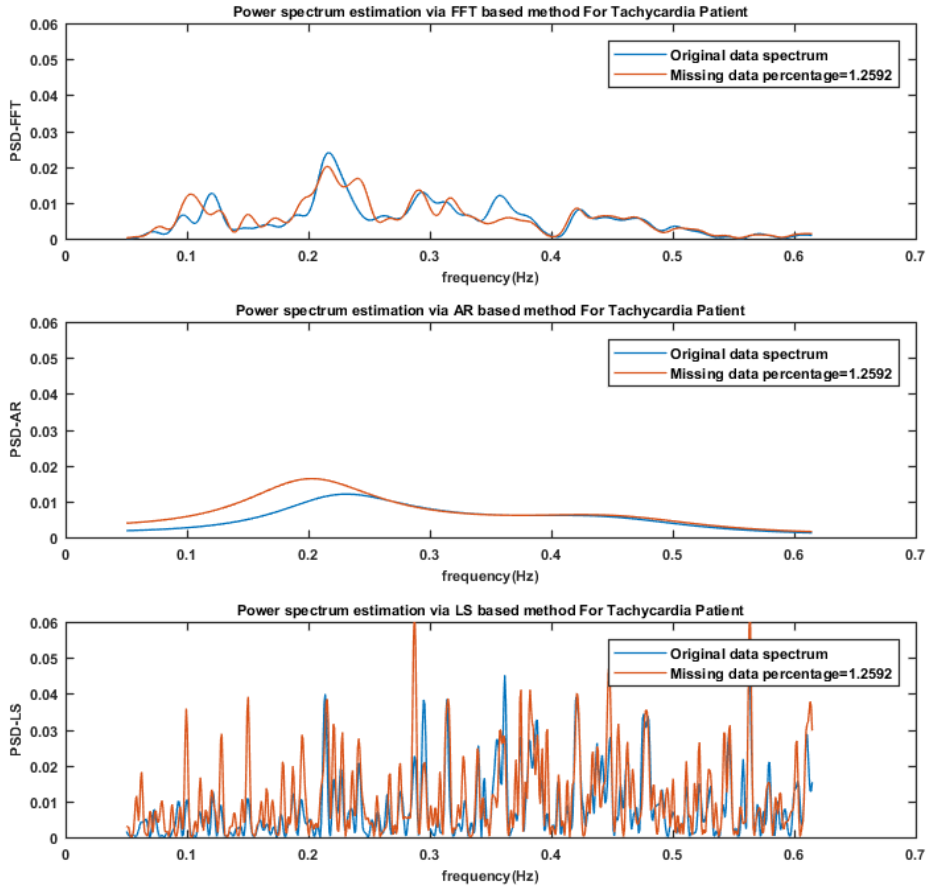

Figure 16. PSD mismatch between non-missing data spectrum and missing data spectrum for tachycardia patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 1.26 percentage missing data.

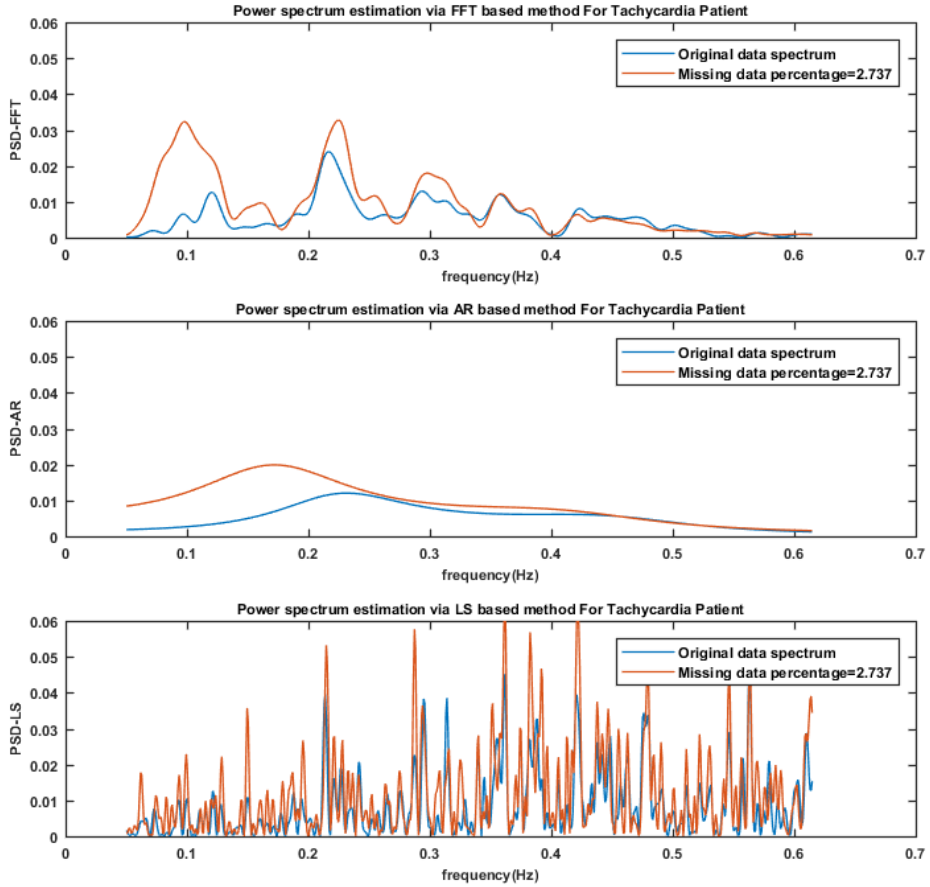

Figure 17. PSD mismatch between non-missing data spectrum and missing data spectrum for tachycardia patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 2.74 percentage missing data.

After that, figure 18 shows the data distribution for tachycardia patient. This figure portraits that, as the increase of missing data percentage, data distribution of error power also climb up for FFT and AR methods respectively in sub-figure 1 and 2. But for the LS method, the data distribution of error power is more compact in the course of missing data percentage compared to the others two methods, it is shown in sub-figure 3. So it is cleared that the effect of missing data was less in LS method like normal and AF patient.

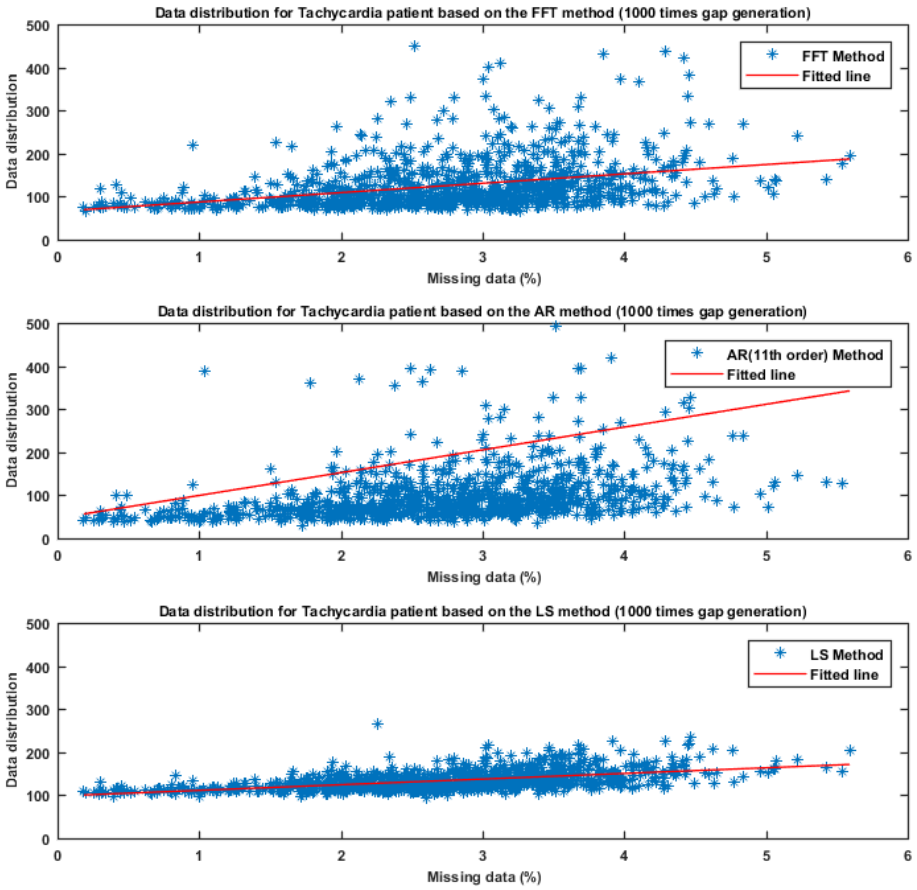

Figure 18. This plot shows the data distribution of error power based on the percentage of missing data. Besides, it describes distribution of random points comparison generated by Monte Carlo simulation for tachycardia patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line.

Furthermore, figure 19 shows the error power percentage between the non-missing data spectrum and missing data spectrum for tachycardia patient through previously described three methods based on the duration of missing data. From sub-figure 1 and 2, median error gradually climbs up as the increase of the missing data percentage for FFT and AR method. On the other hand, from sub-figure 3 the median error power keep almost constant up-to 3.06 percentage of missing data. After that, there is little bit change when missing data goes above 3.06 percentage. Besides, the number of outliers are fewer in LS method compared to the FFT and AR methods in term of the missing data percentage

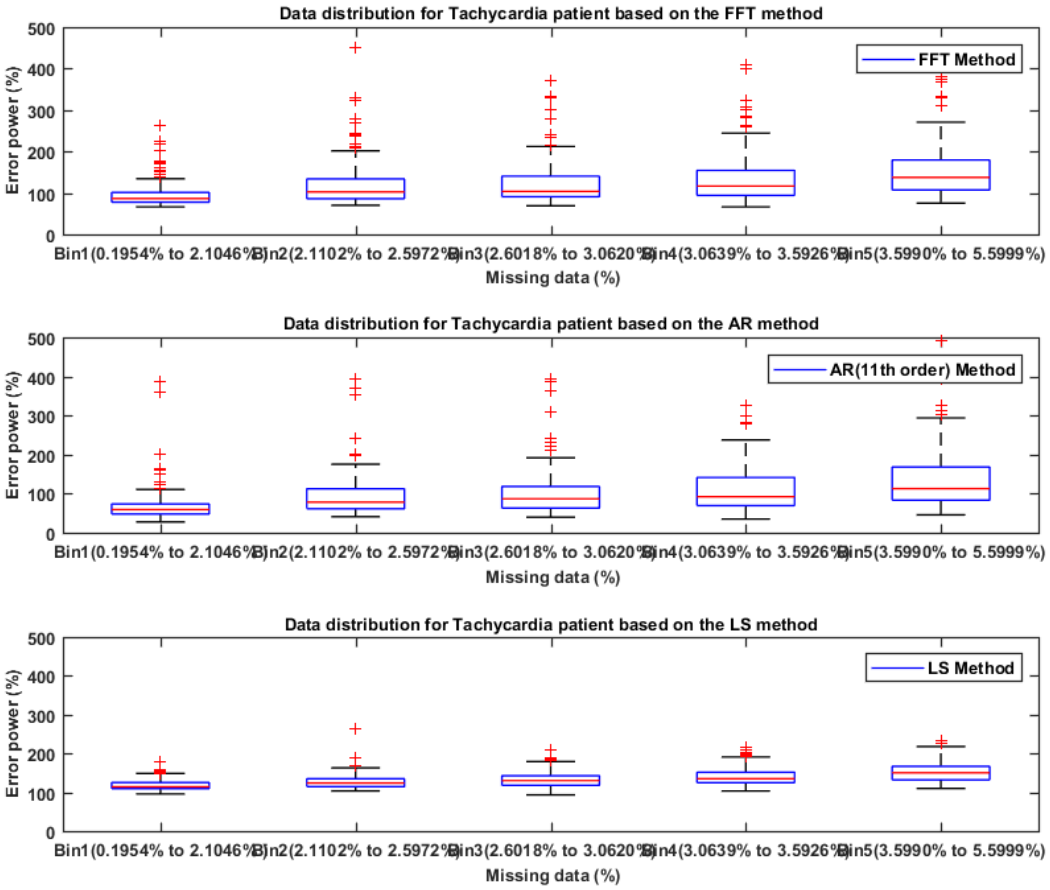

Figure 19. Boxplot of the subfigure 1, 2 and 3 estimated the error power percentage based on the missing data percentage for the tachycardia patient via FFT, AR, and LS method respectively using 1000 times Monte Carlo simulated data. First quartile, median, and third quartile values are portrayed as the bottom, middle and top horizontal line of the boxes. Whiskers are used to represent the most extreme values. And outliers were displayed as crosses. The error power percentage was calculated by equation 4.

4.3.4. Experimental Results for Bradycardia Patient

Finally, 20 to 23 figure represent the experimental results for bradycardia patient, and we used the same technique for bradycardia patient data also. In figure 20 and 21, blue and orange line estimate the power spectrum respectively original data without missing data as the reference values and the percentage of missing data. After that, for a clear perceive, figure 20 and 21 show the PSD mismatch between non-missing data and missing data for FFT , AR and LS based method respectively in sub-figure 1, 2 and 3. From those figures, it has been shown that the maximum peak point and PSD amplitude value are shifted for missing data based on the missing data percentage.

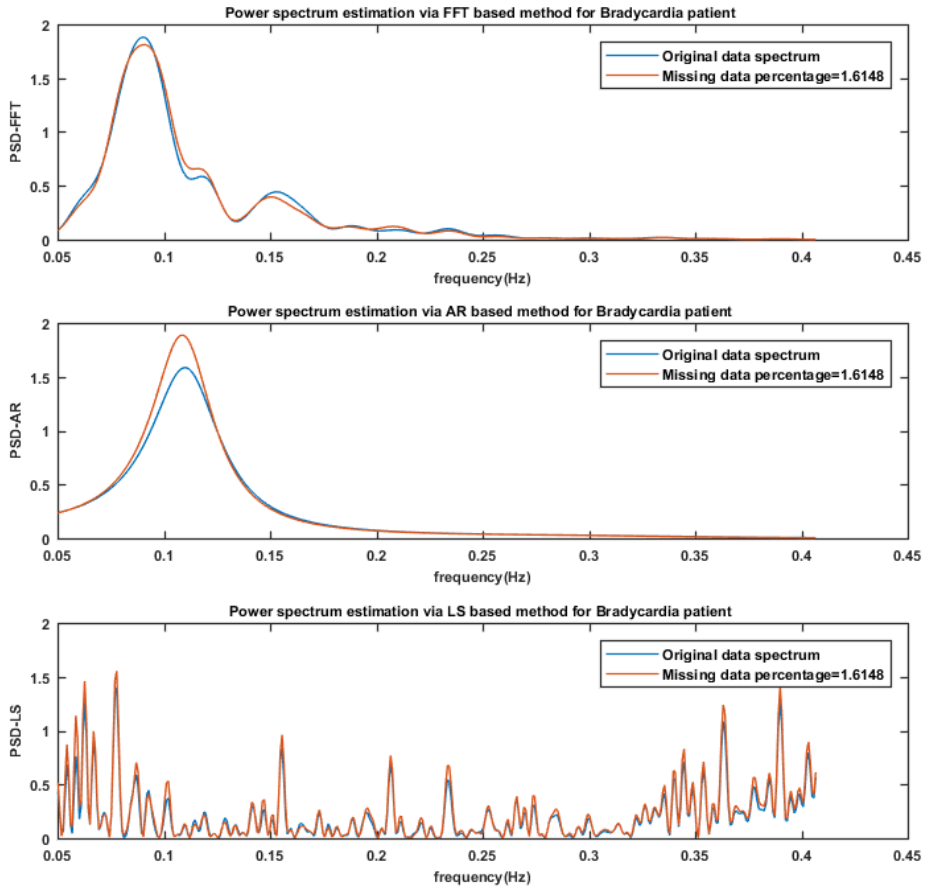

Figure 20. PSD mismatch between non-missing data spectrum and missing data spectrum for bradycardia patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 1.61 percentage missing data.

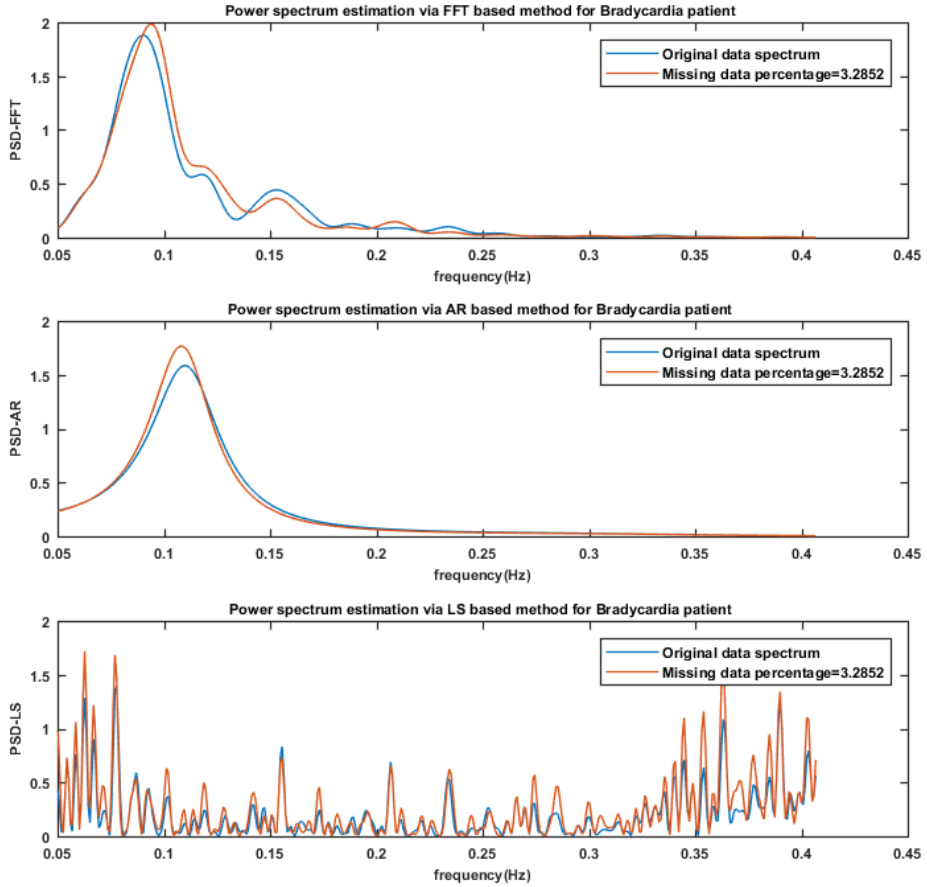

Figure 21. PSD mismatch between non-missing data spectrum and missing data spectrum for bradycardia patient via FFT, AR, LS based method corresponding sub-figure 1, 2 and 3. This figure is generated for 3.29 percentage missing data.

Then, figure 22 shows the data distribution based on the missing data percentage for bradycardia patient. For LS method in sub-figure 3, although the data distribution of error power is higher, it remains more compact over the missing data percentage, whereas data distribution of error power is not much more compact for FFT and AR method, shown in sub-figure 1 and 2 respectively. Now it is visualized that the effect of missing data is less in LS method for bradycardia patient compared to FFT and AR method.

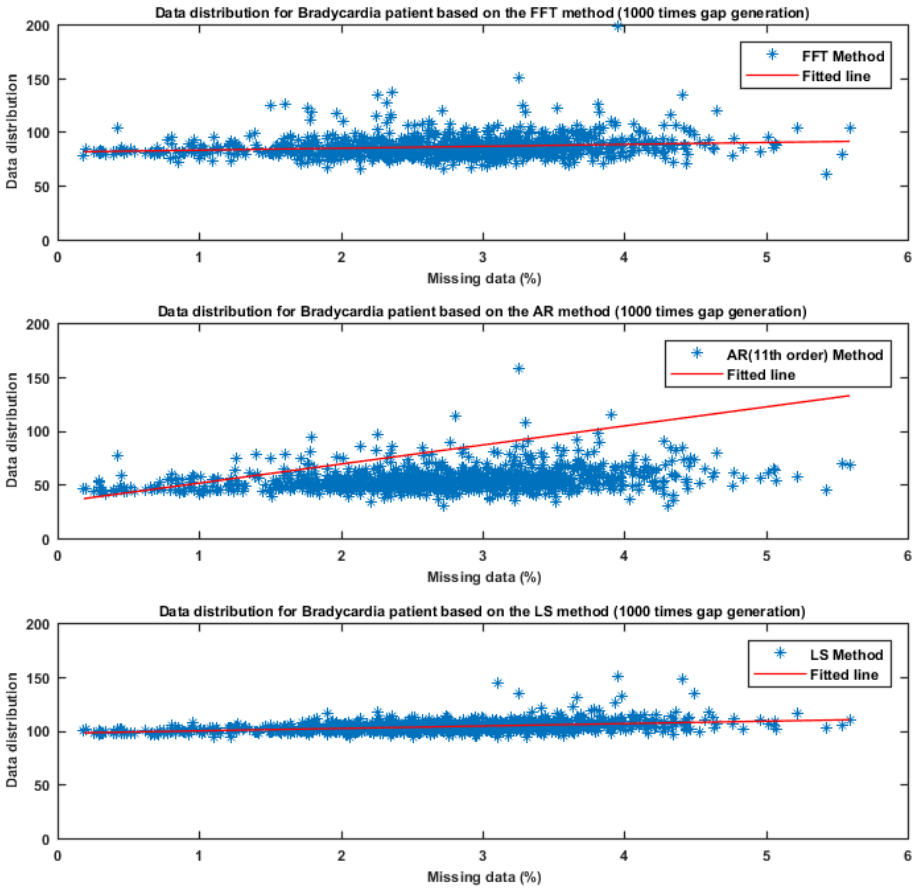

Figure 22. This plot shows the data distribution of error power based on the percentage of missing data. Besides, it describes distribution of random points comparison generated by Monte Carlo simulation for bradycardia patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line.

After that, figure 23 shows the error power percentage between the non-missing data spectrum and missing data spectrum for bradycardia patient via previously described three methods based on the duration of missing data. It is cleared from sub-figure 3 that the median error power lies more or less the same as the percentage of missing data increase although the median error power is higher compared to the other two methods. On the other hand, the median error power remains the same till 2.60 percentage of missing data for FFT and AR methods, showed in sub-figure 1 and 2. After that, the median error minimally increases when the missing data percentage goes above 2.60 percentage. Besides, the number of outliers are more in FFT and AR methods, whereas outliers number are less in the LS method.

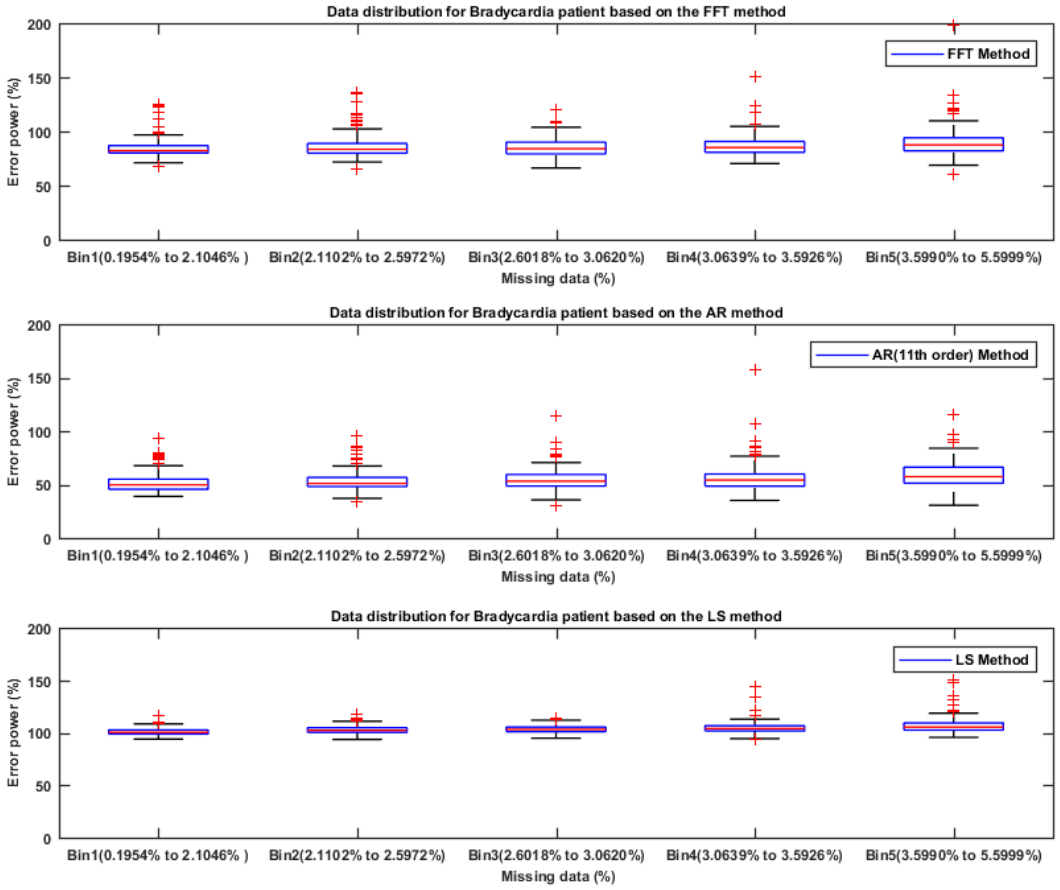

Figure 23. Boxplot of the subfigure 1, 2 and 3 estimated the error power percentage based on the missing data percentage for the bradycardia patient via FFT, AR, and LS method respectively using 1000 times Monte Carlo simulated data. First quartile, median, and third quartile values are portrayed as the bottom, middle and top horizontal line of the boxes. Whiskers are used to represent the most extreme values. And outliers were displayed as crosses. The error power percentage was calculated by the equation 4.

Now figure 24 shows the comparison between those three spectral-based method based on the estimated median value which extracted from the figures [11, 15, 19, 23]. In this figure, sub-figures 1, 2, 3 and 4 represent the normal, AF, tachycardia and bradycardia patient result respectively. From all sub-figures, it is clearly visualized that the AR method shows the significantly lower median value corresponding the missing data percentage compared to the others two methods for normal, AF, tachycardia and bradycardia patient. Finally, the comparison between each method based on the missing data percentage was calculated by Wilcoxon rank sum test. The statistical significance difference between each method result are summarized in table [1, 2, 3 and 4]. All statistical tables have shown in appendix chapter.

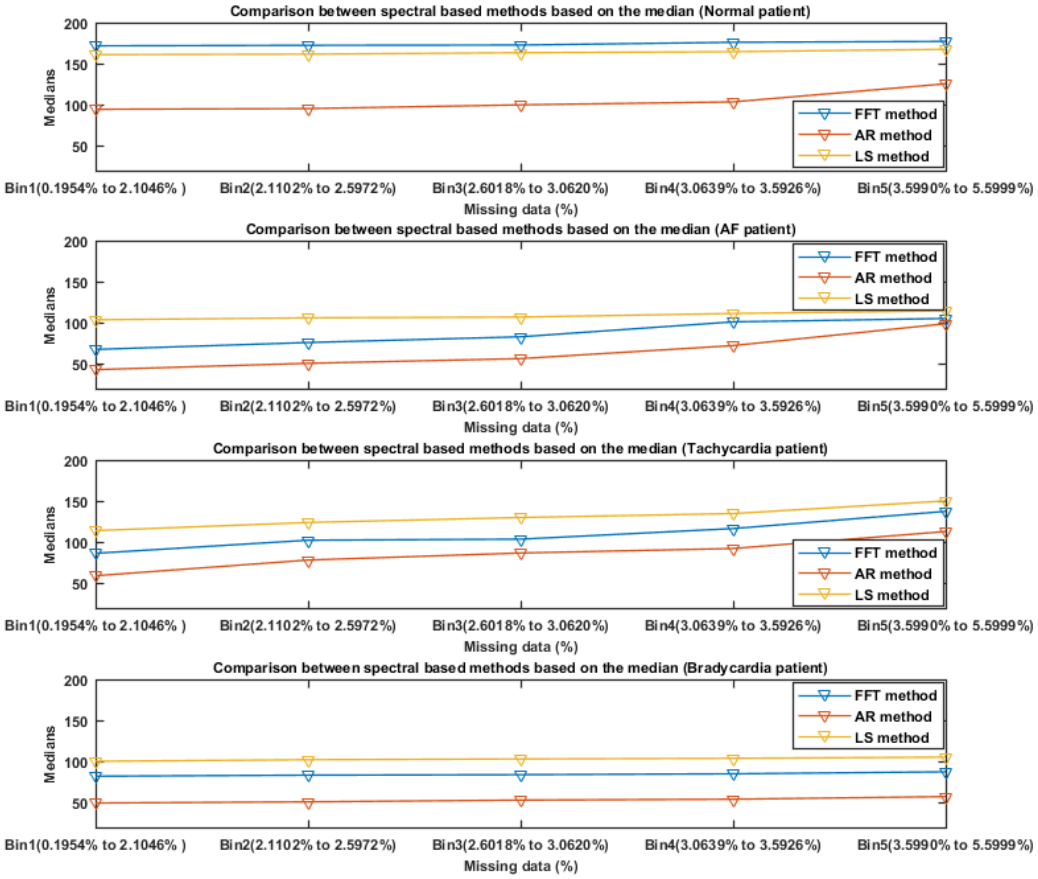

Figure 24. This plot shows the comparison between FFT, AR and LS method based on the estimated median value for normal, AF, tachycardia and bradycardia patient from figure [11, 15, 19, 23]. In figure, the red line, blue line and yellow line represent the AR, FFT and LS method respectively.

4.3.5. Results for Discriminating Different Arrhythmias

In the end, from figure 25 to 29 represent the experimental result of the power spectrum entropy for discriminating between different arrhythmias. In this study, spectrum entropy's result revealed the new dimension to differentiate between arrhythmias.
Figure 25 illustrates that entropy value remains almost the same and much more compact based on the missing data percentage via FFT and AR method for normal patient respectively in sub-figure 1 and 2. On the other hand, in the LS method, entropy value raises and not much more compact with the variation of the missing data percentage in sub-figure 3.

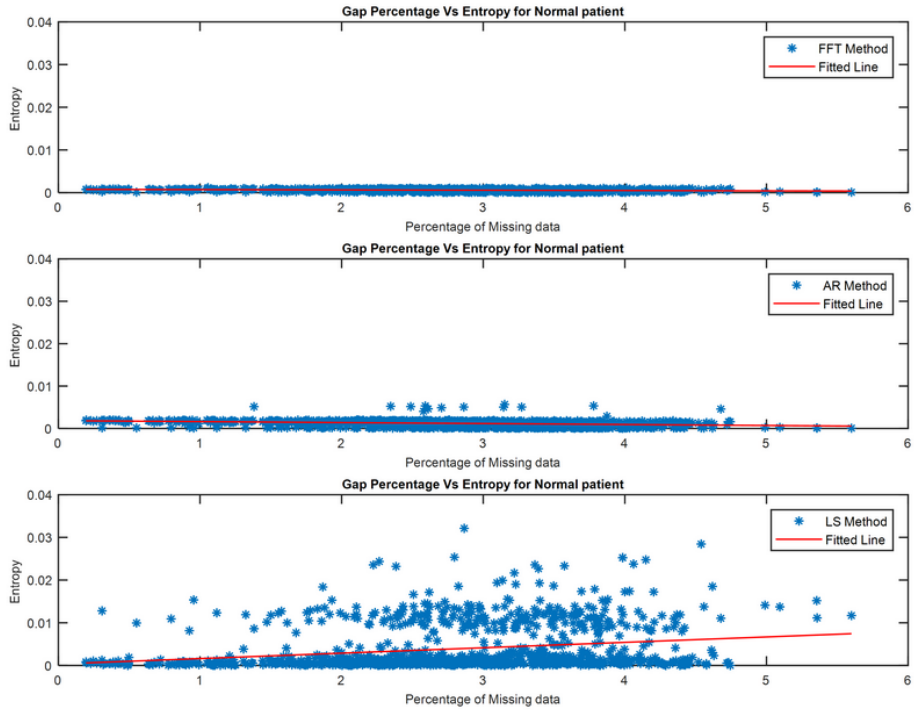

Figure 25. This plot shows the percentage of missing data versus estimated spectral entropy following the analysis of 1000 simulated spectra. Besides, it describes distribution of random points comparison generated by Monte Carlo simulation for normal patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line.

In case of AF patient, figure 26 describes that the entropy value downfall little bit concerning the missing data percentage for FFT and AR method. But in the LS method, it climbs up with the missing data percentage. Figure 27 indicates the entropy value for tachycardia patient. Entropy value in FFT and AR methods are low and remains more or less constant based on the missing data percentage. In contrast, the entropy value in LS method is higher than the other two methods, and it is not much more compact compared to FFT and AR methods.

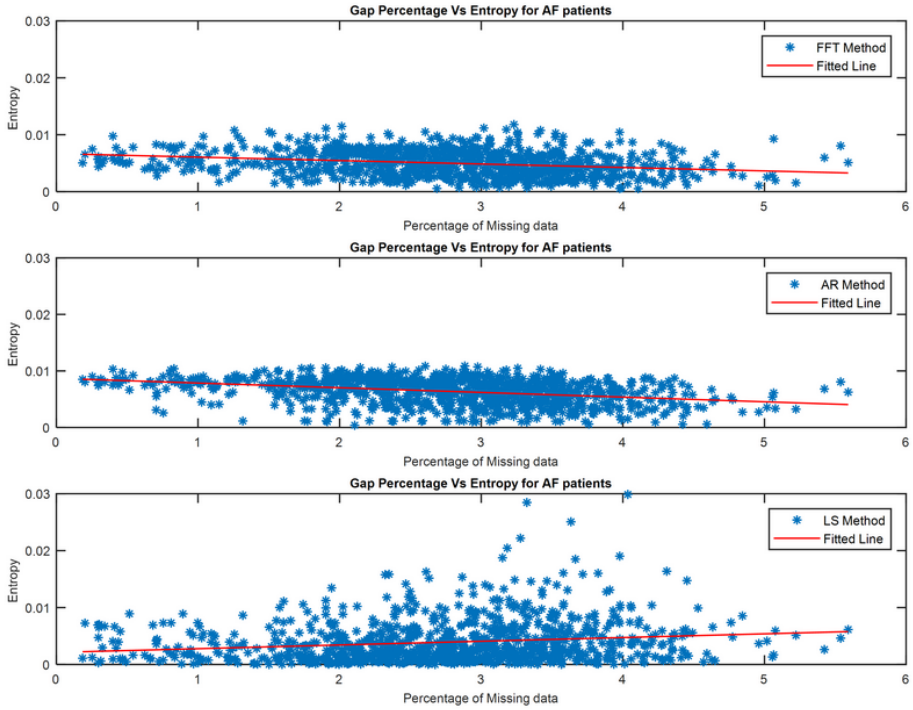

Figure 26. This plot shows the percentage of missing data versus estimated spectral entropy following the analysis of 1000 simulated spectra. Besides, it describes the distribution of random points comparison generated by Monte Carlo simulation for AF patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line.

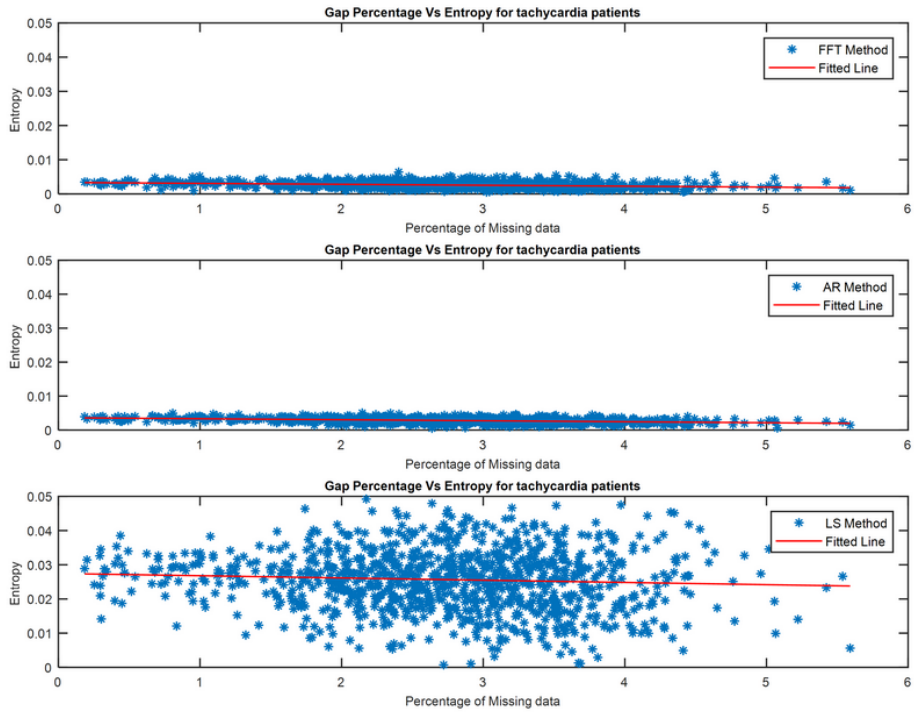

Figure 27. This plot shows the percentage of missing data versus estimated spectral entropy following the analysis of 1000 simulated spectra. Besides, it describes the distribution of random points comparison generated by Monte Carlo simulation for tachycardia patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line.

After that, figure 28 represents the entropy values using above mentioned three spectral based methods for bradycardia patient. In this case, FFT and AR method show the lower entropy, and it remains almost constant corresponding to the missing data percentage. On the contrary, the entropy value in LS method is much higher compared to the others two methods over the missing data percentage.

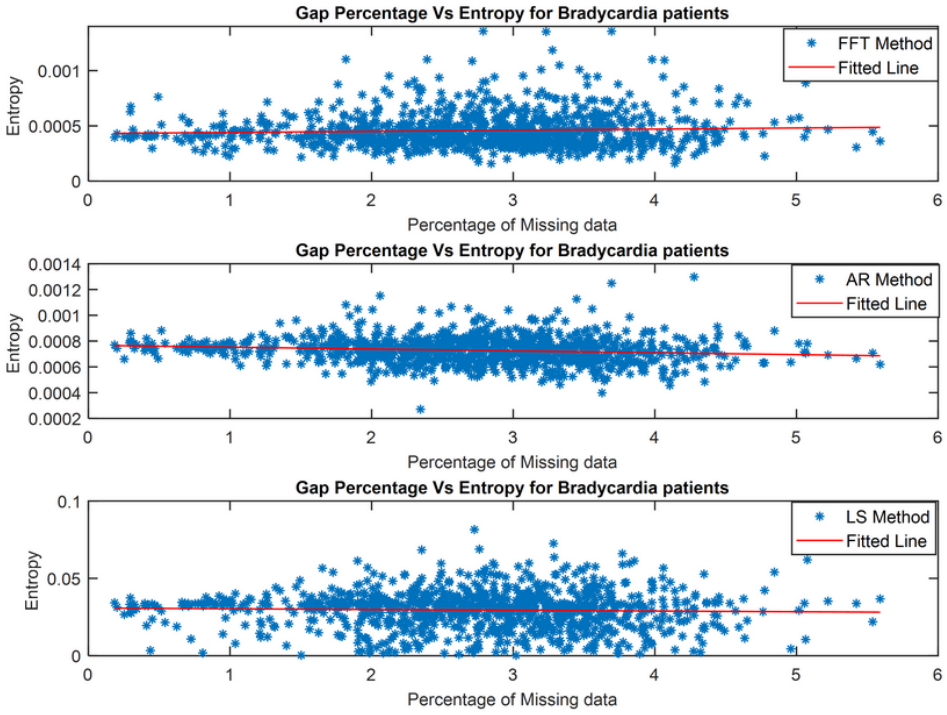

Figure 28. This plot shows the percentage of missing data versus estimated spectral entropy following the analysis of 1000 simulated spectra. Furthermore, It illustrates the distribution of random points comparison generated by Monte Carlo simulation for bradycardia patient using the FFT, AR and LS method respectively sub-figure 1, 2 and 3 and red line indicates the most fitted line. Monte Carlo simulations like this can be used to evaluate the accuracy and reproducibility of those methods.

In all sub-figures of the figure 29 illustrates the difference between arrhythmias based on the missing data percentage using FFT, AR and LS method. Sub-figure 1, 2 and 3 portraits that normal patient has distinctly lower entropy compared to the arrhythmias like AF, tachycardia and bradycardia patient respectively FFT, AR and LS method. Besides, entropy value climbs up than the reference entropy value due to the missing data percentage in FFT method. But reference entropy and median entropy of missing data percentage are almost same in AR and LS method. On the contrary, for arrhythmias cases, entropy value has higher and very much close each other for both methods. Even reference entropy value is so much closer to each other. Furthermore, it is clearly visualized that there is no significant effect of the percentage of missing data on the entropy for arrhythmias cases but the effect is present for the normal patient for both methods. This happens prominently for the LS method which indicates in the sub-figure 3.

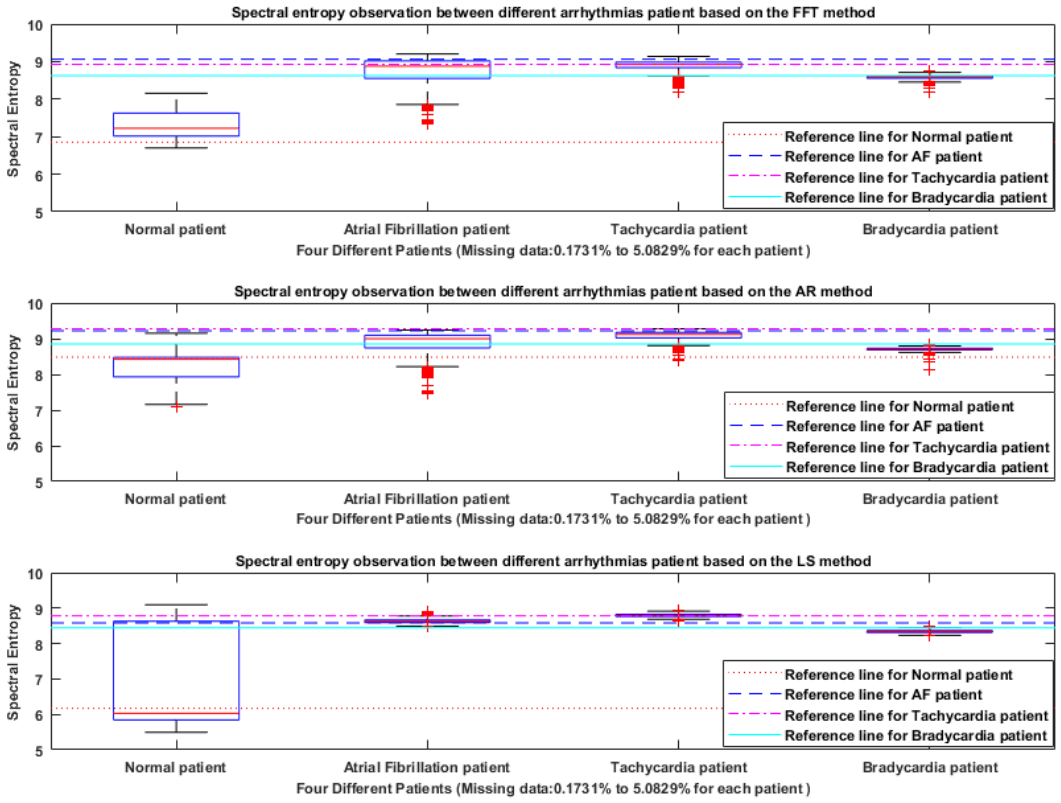

Figure 29. Boxplot of the subfigure 1, 2 and 3 estimated the spectral entropy for different arrhythmias via FFT, AR, and LS method respectively using 1000 times Monte Carlo simulated data for each arrhythmia. First quartile, median, and third quartile values are portrayed as the bottom, middle and top horizontal line of the boxes. Whiskers are used to represent the most extreme values. And outliers were displayed as crosses. In all sub-figures, the red dotted line, blue dashed line, magenta dash-dot line and cyan solid line represent the actual reference entropy when no missing data are present for normal, AF, tachycardia and bradycardia patient.

The spectral entropy of each patient using three spectral based approach was compared by Wilcoxon rank sum test. The statistically significant difference ($p<0.05$) of the entropy values calculated with FFT, AR, and LS based methods. The result is summarized in table 5. The resultant table has shown in appendix chapter.

5. DISCUSSION

5.1. Validity of Result

In this research, we observed the three different spectral-based methods for estimating the PSD with missing data and spectral entropy value for normal, AF, tachycardia and bradycardia patient. The aim of this study was to find out the optimal method to estimate the HRV spectral parameters and to discriminate between arrhythmias based on the spectral entropy value with missing R-R interval data. Besides, we observed the effect of missing data in spectral based three methods. In this chapter, we are trying to address that where the PSD estimation drawbacks and advantages and possible causes of that with missing data via FFT, AR and LS method. Some previous studies described the effect of missing R-R interval data on HRV analysis with different interpolation method [28]. Besides, studies are scarce in the literature which discriminates between arrhythmias based on the spectral entropy. To address the PSD estimation drawbacks and advantages, Fonseca, D.S et al. [29] described that the different interpolation methods distort the spectrum information. In this study, the simulation is occurred in making gap by removing the samples from a full data set. After that, FFT and AR estimation require interpolation the missing intervals. From figures [9, 8, 12, 13, 16, 17, 20 and 21], it is cleared that high frequency components are higher due to interpolation methods and the PSD mismatch is also significant in all comparisons when the missing data percentage is going to be higher. Besides, in the LS method, the PSD mismatch due to missing data was slightly difference from each other. On the other hand, the error power percentage was significant due to the detrending method in parametric methods [29]. Hence, in our study, we have not included the detrending effects for normal, AF, tachycardia and bradycardia patient.

Figure [11, 15, 19 and 23] show the missing data effect of real R-R interval data via FFT, AR and LS method respectively normal, AF, tachycardia and bradycardia patient. In this research, the spline interpolation method is used to make sample evenly for FFT and AR methods. Kim.k et al. [28] explored that the error power percentage increased significantly in the case of LF results due to the spline interpolation method, notably, the duration of missing data section of 5-20 seconds. This time domain duration section is 0.04-0.15 Hz in the frequency domain. Thus, in the process of spline interpolation, missing data in this region are overestimated. On the other hand, they also described that the spline interpolation method shows the small error for small missing data duration of fewer than 60 seconds. These produced low errors are almost similar with simulation results of without missing data.

The observation from those figures, the error power percentage increased based on the missing data percentage for previously described three methods for normal, AF, tachycardia and bradycardia patient. But the LS method shows the comparatively less error for all patients data compared to the FFT and AR methods. Furthermore, we observed that the LS method makes an outstanding result for analyzing data where high percentage missing data, artifact and ectopy are present. Even removing the suspect beat does not effect for analyzing in LS method. Many HRV calculation techniques are evaluated on a non-missing sampled data in time series. For observation the performance of such techniques, the LS method could be returned an optimal resampling

scheme. Scargle [51] mentioned that there is a way to calculate Lomb DFT and its inverse can be made resampling of a time series.

On the contrary, figure 24 reveals that the AR method shows the better performance compared to the FFT and LS method. Because of, the median value for each missing data percentage is lower for normal, AF, tachycardia and bradycardia patient. So from this figure, it is cleared that the AR method comparatively shows the better performance with missing data among others two methods. Besides, the difference is statistically significant based on the median between those spectral based three methods for all cases.

In addition, spectral entropy added a new dimension in this study that gives the best discrimination accuracy to the standard HRV indices. This spectral entropy calculation improves the performance of HRV analysis to discriminate between arrhythmias. Figure 29 estimates the spectral entropy for different arrhythmias with missing data via FFT, AR, and LS method respectively in sub-figure 1, 2 and 3 and different color horizontal lines represent the reference spectral entropy which calculated from non missing data for each case. According to this figure, for a normal patient, the median spectral entropy is more than reference entropy in the FFT method, whereas it is so much closer with reference entropy in the AR method. On the contrary, the median spectral entropy is less than the reference entropy in the LS method. But the difference is little compared to the FFT method. In the case of AF patient, the median spectral entropy is less than the reference entropy in FFT and AR method, whereas the median spectral entropy is more or less same with reference entropy in LS method. For tachycardia patient, the median spectral entropy is almost identical with the reference spectral entropy in FFT and LS method, while the median spectral entropy is less than the reference spectral entropy in the AR method. Now it is clear for bradycardia patient, there is little bit variation for each method. The median spectral entropy is a little bit less than the reference spectral entropy for three methods. Now it is visualized that the normal patient case has distinctly lower entropy, but it is a bit harder to discriminate between the arrhythmias based on the entropy since those are so close to each other. But it is easily distinguishable between normal patient to arrhythmias patient. On the good side, the medians are not entirely overlapped to each other. It could be best possible to classify the correct arrhythmia class. Besides, the percentage of missing data does not have a significant effect on the entropy for the arrhythmias cases like AF, tachycardia and bradycardia patient. But it actually has huge impact on the normal case, especially prominently for the LS method. Besides, The difference between each patient based on the median entropy value is statistically significant for FFT, AR and LS method.

From those three described spectral based methods, one method can be effective for HRV analysis in the frequency domain with missing data. Then, AR method is effective for PSD estimation, but the effect of missing data is less in LS method when the missing data percentage is higher, and the most crucial point revealed by this study is that spectral entropy differentiates between different rhythm which can be used in the diagnosis.

5.2. Future Work

In a future study, nonlinear HRV parameter will be observed with missing data for various diseases patients. Besides, skewness and kurtosis can be used to characterizing the various diseases patients from normal patients with different percentage missing data via FFT, AR and LS method through further studies.

6. SUMMARY

The role of the autonomic nervous system (ANS) can be described by the standard heart rate variability (HRV) method. HRV analysis can be occurred based on the time-domain and frequency-domain. Robust spectral HRV features calculation with missing ECG data has been studied from mid 90th. There are not many studies which calculate the HRV spectral features with missing ECG signal for different kind of patients data. Our research focused on two things. Firstly, it focused on the comparison between three spectral methods performance based on the missing data percentage with normal and three different types of patients data and also observed the effect of missing data among those three spectral methods. Secondly, It's described that the spectral entropy measurement from the HRV index for discriminating between arrhythmias. In this research, one subject has experimented for each normal, AF, tachycardia and bradycardia patient. We used 5 minutes of ECG data for each case and generated 1000 times Monte Carlo simulation for making the artificial gap in R-R tachogram. After removing the baseline-wandering, R peak detection is the basic signal extraction from ECG signal. This dataset was obtained from the MIT/BIH database.
Basically, using the artificial R-R tachogram, FFT and AR methods have been used for estimating the PSD of resampled tachogram with completely known components and LS method does not require interpolation or resampling like FFT and AR methods. This research presents that the artificial gap generation from R-R tachogram and that the calculation of HRV through FFT and AR methods generates significant errors due to the resampling step. The LS method is performed to reduce these errors even in a large amount of missing data. By observing the result, it is cleared that LS method shows the lower variance estimate of the frequency components compared to the FFT and AR method for normal, AF, tachycardia and bradycardia patient based on the missing data percentage. This lower variance estimation indicates that the LS method for calculating spectral HRV are more robust and accurate corresponding to the reference signal. On the other hand, the AR method shows better performance when we compare the PSD estimation based on the median among those three methods.

In addition, this study revealed the most crucial point that spectral entropy measurement using HRV index discriminates between arrhythmias via described three methods in chapter 3. From the comparison figure 29, it is revealed that reference spectral entropy value using HRV index is much lower for the normal patient using FFT, AR and LS method, whereas reference spectral entropies value for arrhythmias cases are closer each other. But it is cleared that the reference spectral entropy for bradycardia patient has distinguishable lower value compared to AF and tachycardia patient. Besides, the reference spectral entropies value of AF and tachycardia are much closer in both methods, but it is differentiable.

To sum up our study, the first part of our experiment revealed that the LS method provides a more robust and accurate result for estimating spectral HRV with missing data corresponding to the reference data. On the other hand, AR shows the better performance to determine the PSD with a certain percentage of missing data. The results of our last part present that entropy measures of ECG signal have the importance sides to identify different arrhythmias via FFT, AR and LS method. Besides, the most important point that the combination of features that give more discrimination accuracies include spectral entropy measures for discriminating the different arrhythmias.

7. REFERENCES

[1] İşler Y. & Kuntalp M. (2007) Combining classical hrv indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. Computers in biology and medicine 37, pp. 1502–1510.

[2] Clifford G.D. & Tarassenko L. (2005) Quantifying errors in spectral estimates of hrv due to beat replacement and resampling. IEEE transactions on biomedical engineering 52, pp. 630–638.

[3] Fonseca D., Netto A., Ferreira R. & de Sá A.M. (2013) Lomb-scargle periodogram applied to heart rate variability study. In: Biosignals and Biorobotics Conference (BRC), 2013 ISSNIP, IEEE, pp. 1–4.

[4] Clayton R., Lord S., McComb J. & Murray A. (1997) Comparison of autoregressive and fourier transform based techniques for estimating rr interval spectra. In: Computers in Cardiology 1997, IEEE, pp. 379–382.

[5] Malik M., Bigger J.T., Camm A.J., Kleiger R.E., Malliani A., Moss A.J. & Schwartz P.J. (1996) Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European heart journal 17, pp. 354–381.

[6] Birkett C., Kienzle M. & Myers G. (1991) Interpolation over ectopic beats increases low frequency power in heart rate variability spectra. In: Computers in Cardiology 1991, Proceedings., IEEE, pp. 257–259.

[7] Piegl L. (1991) On nurbs: a survey. IEEE Computer Graphics and Applications 11, pp. 55–71.

[8] Lippman N., Stein K.M. & Lerman B.B. (1994) Comparison of methods for removal of ectopy in measurement of heart rate variability. American Journal of Physiology-Heart and Circulatory Physiology 267, pp. H411–H418.

[9] Lomb N.R. (1976) Least-squares frequency analysis of unequally spaced data. Astrophysics and space science 39, pp. 447–462.

[10] Scargle J.D. (1982) Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. The Astrophysical Journal 263, pp. 835–853.

[11] Press W.H., Teukolsky S.A., Vetterling W.T. & Flannery B.P. (2007) Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press.

[12] Moody G.B. (1993) Spectral analysis of heart rate without resampling. In: Computers in Cardiology 1993, Proceedings., IEEE, pp. 715–718.

[13] Abeysekera S., Abeyratne U.R. & Goh S. (2000) Spectral information changes in obtaining heart rate variability from tachometer rr interval signals. Critical Reviews™ in Biomedical Engineering 28.

[14] Zheng-You H., Xiaoqing C. & Guoming L. (2006) Wavelet entropy measure definition and its application for transmission line fault detection and identification;(part i: Definition and methodology). In: Power System Technology, 2006. PowerCon 2006. International Conference on, IEEE, pp. 1–6.

[15] Commons W. (2007), Sinusrhythmlabels. URL: https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.svg.

[16] Ahmed W. & Khalid S. (2016) Ecg signal processing for recognition of cardiovascular diseases: A survey. In: Innovative Computing Technology (INTECH), 2016 Sixth International Conference on, IEEE, pp. 677–682.

[17] Alikhani I. ECG-Derived Respiration Estimation During Daily Activites. Faculty of Information Technology and Electrical Engineering, University of Oulu,Oulu, Finland. Master's thesis.

[18] Gustafson P. & Hagblad J. (2007) Miniaturised multi channel wireless ecg connected to graphical computer interface using bluetooth. The Department of Computer Science and Electronics, Mälardalen University .

[19] Malmivuo J., Plonsey R. et al. (1995) Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields. Oxford University Press, USA.

[20] Kumar M., Umesh K., Pandiaraja G., Thomas S. & Venkatesh M. (2014) A research work on identification of cardiac disorders using labview .

[21] Spodick D.H., Raju P., Bishop R.L. & Rifkin R.D. (1992) Operational definition of normal sinus heart rate. American Journal of Cardiology 69, pp. 1245–1246.

[22] Krahn A.D., Yee R., Klein G.J. & Morillo C. (1995) Inappropriate sinus tachycardia: evaluation and therapy. Journal of cardiovascular electrophysiology 6, pp. 1124–1128.

[23] Adan V. & Crown L.A. (2003) Diagnosis and treatment of sick sinus syndrome. American family physician 67, pp. 1725–1732.

[24] Billeci L., Chiarugi F., Costi M., Lombardi D., Varanini M. & SpA C. (2017) Detection of af and other rhythms using rr variability and ecg spectral measures. Computing 44, p. 1.

[25] Cooper J.C. (2005) The poisson and exponential distributions. Mathematical Spectrum 37, pp. 123–125.

[26] Rastogi N. & Mehra R. (2013) Analysis of savitzky-golay filter for baseline wander cancellation in ecg using wavelets. Int. J. Eng. Sci. Emerg. Technol 6, pp. 15–23.

[27] Wen F. & He F.t. (2011) An efficient method of addressing ectopic beats: new insight into data preprocessing of heart rate variability analysis. Journal of Zhejiang University SCIENCE B 12, pp. 976–982.

[28] Kim K.K., Kim J.S., Lim Y.G. & Park K.S. (2009) The effect of missing rr-interval data on heart rate variability analysis in the frequency domain. Physiological measurement 30, p. 1039.

[29] Fonseca D., Netto A., Ferreira R. & de Sá A.M. (2013) Lomb-scargle periodogram applied to heart rate variability study. In: Biosignals and Biorobotics Conference (BRC), 2013 ISSNIP, IEEE, pp. 1–4.

[30] Proakis J.G. (2001) Digital signal processing: principles algorithms and applications. Pearson Education India.

[31] Welch P. (1967) The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on audio and electroacoustics 15, pp. 70–73.

[32] Solomon Jr O. (1991) Psd computations using welch's method. NASA STI/Recon Technical Report N 92.

[33] Sassi R., Mainardi L., Maison Blanche P. & Cerutti S. (2005) Estimation of spectral parameters of residual ecg signal during atrial fibrillation using autoregressive models. Folia Cardiologica 12, pp. 108–110.

[34] Brockwell P. & Dahlhaus R. (2004) Generalized levinson–durbin and burg algorithms. Journal of Econometrics 118, pp. 129–149.

[35] Νεοφύτου N. (2012) ECG event detection & recognition using time-frequency analysis. Ph.D. thesis.

[36] VanderPlas J.T. (2018) Power Spectral Density of Unevenly Sampled Heart Rate Data * . The Astrophysical Journal Supplement Series, 236:16 (28pp), 2018 May 15, pp. 236–254.

[37] Laguna P., Moody G. & Mark R. (1995) Power spectral density of unevenly sampled heart rate data. In: Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference, vol. 1, IEEE, vol. 1, pp. 157–158.

[38] KAMATH C. (2013) Quantification of electrocardiogram rhythmicity to detect life threatening cardiac arrhythmias using spectral entropy. Journal of Engineering Science and Technology 8, pp. 588–602.

[39] Hu M., Li J., Li G., Tang X. & Ding Q. (2006) Classification of normal and hypoxia eeg based on approximate entropy and welch power-spectral-density. In: Neural Networks, 2006. IJCNN'06. International Joint Conference on, IEEE, pp. 3218–3222.

[40] Zhang A., Yang B. & Huang L. (2008) Feature extraction of eeg signals using power spectral entropy. In: 2008 International Conference on BioMedical Engineering and Informatics, IEEE, pp. 435–439.

[41] Inouye T., Shinosaki K., Sakamoto H., Toi S., Ukai S., Iyama A., Katsuda Y. & Hirano M. (1991) Quantification of eeg irregularity by use of the entropy of the power spectrum. Electroencephalography and clinical Neurophysiology 79, pp. 204–210.

[42] Inouye T., Shinosaki K., Sakamoto H., Toi S., Ukai S., Iyama A., Katsuda Y. & Hirano M. (1992) Abnormality of background eeg determined by the entropy of power spectra in epileptic patients. Electroencephalography and clinical Neurophysiology 82, pp. 203–207.

[43] Yoon Y.G., Kim T.H., Jeong D.W. & Park S.H. (2011) Monitoring the depth of anesthesia from rat eeg using modified shannon entropy analysis. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, IEEE, pp. 4386–4389.

[44] Zhou P., Zong-Xia M., Chun-Lan H. & Huang Y.X. (2017) Power spectral entropy in the ecg of patients suffered from nocturnal frontal lobe epilepsy. Journal of Pharmaceutical and Biomedical Sciences 7.

[45] Yu L., Sun X. & Zhang K. (2011) Driving distraction analysis by ecg signals: an entropy analysis. In: International Conference on Internationalization, Design and Global Development, Springer, pp. 258–264.

[46] Misra H., Ikbal S., Bourlard H. & Hermansky H. (2004) Spectral entropy based feature for robust asr. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), EPFL-CONF-83132.

[47] Zheng-You H., Xiaoqing C. & Guoming L. (2006) Wavelet entropy measure definition and its application for transmission line fault detection and identification;(part i: Definition and methodology). In: Power System Technology, 2006. PowerCon 2006. International Conference on, IEEE, pp. 1–6.

[48] Mamaghani B., Sterling M., Gruendike D., Hamer M. & Ghoraani B. (2014) Entropy and frequency analysis of new electrocardiogram lead placement .

[49] Misra H., Ikbal S., Bourlard H. & Hermansky H. (2004) Spectral entropy based feature for robust asr. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), EPFL-CONF-83132.

[50] Zheng-You H., Xiaoqing C. & Guoming L. (2006) Wavelet entropy measure definition and its application for transmission line fault detection and identification;(part i: Definition and methodology). In: Power System Technology, 2006. PowerCon 2006. International Conference on, IEEE, pp. 1–6.

[51] Scargle J.D. (1989) Studies in astronomical time series analysis. iii-fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. The Astrophysical Journal 343, pp. 874–887.

8. APPENDICES

Table 1. Statistical paired ranksum test comparison table of median values between three spectral based methods for a normal patient. The statistical the significant difference ($p<0.05$) of the median values obtained with normal patient data in comparison to others method is indicated (asterisk*)

percentage of missing data	Methods	AR	LS
Bin1 (0.1954 to 2.1046)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin2 (2.1102 to 2.5972)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin3 (2.6018 to 3.0620)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin4 (3.0639 to 3.5926)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin5 (3.5990 to 5.5999)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*

Table 2. Statistical paired ranksum test comparison table of median values between three spectral based methods for AF patient. The statistical the significant difference ($p<0.05$) of the median values obtained with AF patient data in comparison to others method is indicated (asterisk*)

percentage of missing data	Methods	AR	LS
Bin1 (0.1954 to 2.1046)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin2 (2.1102 to 2.5972)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin3 (2.6018 to 3.0620)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin4 (3.0639 to 3.5926)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin5 (3.5990 to 5.5999)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*

Table 3. Statistical paired ranksum test comparison table of median values between three spectral based methods for tachycardia patient. The statistical the significant difference (p<0.05) of the median values obtained with AF patient data in comparison to others method is indicated (asterisk*)

percentage of missing data	Methods	AR	LS
Bin1 (0.1954 to 2.1046)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin2 (2.1102 to 2.5972)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin3 (2.6018 to 3.0620)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin4 (3.0639 to 3.5926)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin5 (3.5990 to 5.5999)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*

Table 4. Statistical paired ranksum test comparison table of median values between three spectral based methods for bradycardia patient. The statistical the significant difference (p<0.05) of the median values obtained with AF patient data in comparison to others method is indicated (asterisk*)

percentage of missing data	Methods	AR	LS
Bin1 (0.1954 to 2.1046)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin2 (2.1102 to 2.5972)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin3 (2.6018 to 3.0620)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin4 (3.0639 to 3.5926)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*
Bin5 (3.5990 to 5.5999)	FFT	p<0.05*	p<0.05*
	AR	-	p<0.05*

Table 5. Statistical paired ranksum comparison table of entropy values between Normal, AF, tachycardia and Bradycardia patient via three spectral based methods. The statistically significant difference ($p<0.05$) of the entropy values obtained with FFT, AR and LS based methods in comparison to others patients is indicated (asterisk*)

Spectral method	Patients	Atrial fibrillation	Tachycardia	Bradycardia
FFT	Normal	p<0.05*	p<0.05*	p<0.05*
	Atrial fibrillation	-	p<0.05*	p<0.05*
	Tachycardia	-	-	p<0.05*
AR	Normal	p<0.05*	p<0.05*	p<0.05*
	Atrial fibrillation	-	p<0.05*	p<0.05*
	Tachycardia	-	-	p<0.05*
LS	Normal	p<0.05*	p<0.05*	p<0.05*
	Atrial fibrillation	-	p<0.05*	p<0.05*
	Tachycardia	-	-	p<0.05*